



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> C12N 15/48, 15/08, 7/02, C07K 14/15, C12N 9/12, 9/22, C12Q 1/70, C07K 16/10, G01N 33/569, A61K 39/21, 39/42, 48/00	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 98/23755</b>  <b>(43) International Publication Date:</b> 4 June 1998 (04.06.98)
<b>(21) International Application Number:</b> PCT/IB97/01482 <b>(22) International Filing Date:</b> 26 November 1997 (26.11.97)  <b>(30) Priority Data:</b> 08/756,429 26 November 1996 (26.11.96) US  <b>(71) Applicant:</b> BIO MERIEUX [FR/FR]; Chemin de l'Orme, F-69280 Marcy l'Etoile (FR).  <b>(72) Inventors:</b> PERRON, Hervé; 134, rue du Docteur E. Lo- card, F-69005 Lyon (FR). BESEME, Frédéric; 39, rue de la Noyera, F-38090 Villefontaine (FR). BEDIN, Frédéric; 6, rue Gaspard André, F-69002 Lyon (FR). PARANHOS-BACCALA, Glaucia; 75, cours Gambetta, F-69003 Lyon (FR). KOMURIAN-PRADEL, Florence; Chemin Vial, F-69450 Saint-Cyr-au-Mont d'Or (FR). JOLIVET-REYNAUD, Colette; 16, avenue des Colonnes, F-69500 Bron (FR). MANDRAND, Bernhard; 21, rue de la Doua, F-69100 Villeurbanne (FR).  <b>(74) Agent:</b> CABINET GERMAIN & MAUREAU; Boîte postale 6153, F-69466 Lyon Cedex 06 (FR).		<b>(81) Designated States:</b> CA, JP, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
<b>(54) Title:</b> VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOS- TIC, PROPHYLACTIC AND THERAPEUTIC PURPOSES  <b>(57) Abstract</b>  The invention relates to a nucleic material, in the isolated or purified state, comprising a nucleotide sequence selected from the group including sequences SEQ ID NO:93, SEQ ID NO:94, their complementary sequences and their equivalent sequences, in particular nucleotide sequences displaying, for any succession of 100 contiguous monomers, at least 50 % and preferably at least 60 % homology with said sequences SEQ ID NO:93, SEQ ID NO:94 and their complementary sequences, excluding HSERV-9 sequence.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS ASSOCIATED WITH  
MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC AND  
THERAPEUTIC PURPOSES

5           Multiple sclerosis (MS) is a demyelinating disease of the central nervous system (CNS) the cause of which remains as yet unknown.

          "Multiple sclerosis (MS) is the most common neurological disease of young adults with a prevalence in  
10 Europe and North America of between 20 and 200 per 100,000. It is characterized clinically by a relapsing/remitting or chronic progressive course, frequently leading to severe disability. Current knowledge suggests that MS is associated with autoimmunity, that  
15 genetic background has an important influence and that "infectious" agent(s) may be involved. Indeed, many viruses have been proposed as possible candidates but as yet, none of them has been shown to play an aetiological role.

20           Many studies have supported the hypothesis of a viral aetiology of the disease, but none of the known viruses tested has proved to be the causal agent sought: a review of the viruses sought for several years in MS has been compiled by E. Norrby (1) and R.T. Johnson (2).

25           The discovery of pathogenic retroviruses in man (HTLVs and HIVs) was followed by great interest in their ability to impair the immune system and to provoke central nervous system inflammation and/or degeneration. In the case of HTLV-1, its association with a chronic  
30 inflammatory demyelinating disease in man (48) led to extensive investigations to search for an HTLV1-like retrovirus in MS patients. However, despite initial claims, the presence of HTLV-1 or HTLV-like retroviruses was not confirmed.

Recently, a retrovirus different from the known human retroviruses has been isolated in patients suffering from MS (3, 4, and 5).

In 1989, the authors described the production of  
5 extracellular virions, associated with reverse  
transcriptase (RT) activity, by a culture of  
leptomeningeal cells (LM7) obtained from the cerebrospinal  
fluid of a patient with MS (3). This was followed by  
similar findings in monocyte cultures from a series of MS  
10 patients (5). Neither viral particles nor viral RT-  
activity were found in control individuals. Furthermore,  
the authors were able to transfer the LM7 virus to non-  
infected leptomeningeal cells *in vitro* (26). The molecular  
characterization of the "LM7" retrovirus was a  
15 prerequisite for further evaluation of its possible role  
in MS. Considerable difficulties arose from the absence of  
continuously productive retroviral cultures and from the  
low levels of expression in the few transient cultures.  
The strategy described here focused on RNA from  
20 extracellular virions, in order to avoid non-specific  
detection of cellular RNA and of endogenous elements from  
contaminating human DNA. A specific retroviral sequence  
associated with virions produced by cell cultures from  
several MS patients has been identified. The entire  
25 sequence of this novel retroviral genome is currently  
being obtained using RT-PCR on RNA from extracellular  
virions. The retrovirus previously called "LM7 virus"  
corresponds to an oncovirus and is now designated MSRV  
(Multiple Sclerosis-associated RetroVirus).

30 The authors were also able to show that this  
retrovirus could be transmitted *in vitro*, that patients  
suffering from MS produced antibodies capable of  
recognizing proteins associated with the infection of  
leptomeningeal cells by this retrovirus, and that the  
35 expression of the latter could be strongly stimulated by  
the immediate-early genes of some herpesviruses (6).



All these results point to the role in MS of at least one unknown retrovirus or of a virus having reverse transcriptase activity which is detectable according to the method published by H. Perron (3) and qualified as "LM7-like RT" activity. The content of the publication identified by (3) is incorporated in the present description by reference.

Recently, the Applicant's studies have enabled two continuous cell lines infected with natural isolates originating from two different patients suffering from MS to be obtained by a culture method as described in the document WO-A-93/20188, the content of which is incorporated in the present description by reference. These two lines, derived from human choroid plexus cells, designated LM7PC and PLI-2, were deposited with the ECACC on 22nd July 1992 and 8th January 1993, respectively, under numbers 92072201 and 93010817, in accordance with the provisions of the Budapest Treaty. Moreover, the viral isolates possessing LM7-like RT activity were also deposited with the ECACC under the overall designation of "strains". The "strain" or isolate harboured by the PLI-2 line, designated POL-2, was deposited with the ECACC on 22nd July 1992 under No. V92072202. The "strain" or isolate harboured by the LM7PC line, designated MS7PG, was deposited with the ECACC on 8th January 1993 under No. V93010816.

Starting from the cultures and isolates mentioned above, characterized by biological and morphological criteria, the next step was to endeavour to characterize the nucleic acid material associated with the viral particles produced in these cultures.

The portions of the genome which have already been characterized have been used to develop tests for molecular detection of the viral genome and immunoserological tests, using the amino acid sequences encoded by the nucleotide sequences of the viral genome,

in order to detect the immune response directed against epitopes associated with the infection and/or viral expression.

These tools have already enabled an association to be confirmed between MS and the expression of the sequences identified in the patents cited later. However, the viral system discovered by the Applicant is related to a complex retroviral system. In effect, the sequences to be found encapsidated in the extracellular viral particles produced by the different cultures of cells of patients suffering from MS show clearly that there is coencapsidation of retroviral genomes which are related but different from the "wild-type" retroviral genome which produces the infective viral particles. This phenomenon has been observed between replicative retroviruses and endogenous retroviruses belonging to the same family, or even heterologous retroviruses. The notion of endogenous retroviruses is very important in the context of our discovery since, in the case of MSRV-1, it has been observed that endogenous retroviral sequences comprising sequences homologous to the MSRV-1 genome exist in normal human DNA. The existence of endogenous retroviral elements (ERV) related to MSRV-1 by all or part of their genome explains the fact that the expression of the MSRV-1 retrovirus in human cells is able to interact with closely related endogenous sequences. These interactions are to be found in the case of pathogenic and/or infectious endogenous retroviruses (for example some ecotropic strains of the murine leukaemia virus), and in the case of exogenous retroviruses whose nucleotide sequence may be found partially or wholly, in the form of ERVs, in the host animal's genome (e.g. mouse exogenous mammary tumor virus transmitted via the milk). These interactions consist mainly of (i) a trans-activation or coactivation of ERVs by the replicative retrovirus (ii) and "illegitimate" encapsidation of RNAs related to ERVs, or

of ERVs - or even of cellular RNAs - simply possessing compatible encapsidation sequences, in the retroviral particles produced by the expression of the replicative strain, which are sometimes transmissible and sometimes  
5 with a pathogenicity of their own, and (iii) more or less substantial recombinations between the coencapsidated genomes, in particular in the phases of reverse transcription, which lead to the formation of hybrid genomes, which are sometimes transmissible and sometimes  
10 with a pathogenicity of their own.

Thus, (i) different sequences related to MSRV-1 have been found in the purified viral particles; (ii) molecular analysis of the different regions of the MSRV-1 retroviral genome should be carried out by systematically  
15 analyzing the coencapsidated, interfering and/or recombined sequences which are generated by the infection and/or expression of MSRV-1; furthermore, some clones may have defective sequence portions produced by the retroviral replication and template errors and/or errors  
20 of transcription of the reverse transcriptase; (iii) the families of sequences related to the same retroviral genomic region provide the means for an overall diagnostic detection which may be optimized by the identification of invariable regions among the clones expressed, and by the  
25 identification of reading frames responsible for the production of antigenic and/or pathogenic polypeptides which may be produced only by a portion, or even by just one, of the clones expressed, and, under these conditions, the systematic analysis of the clones expressed in the  
30 region of a given gene enables the frequency of variation and/or of recombination of the MSRV-1 genome in this region to be evaluated and the optimal sequences for the applications, in particular diagnostic applications, to be defined; (iv) the pathology caused by a retrovirus such as  
35 MSRV-1 may be a direct effect of its expression and of the proteins or peptides produced as a result thereof, but

also an effect of the activation, the encapsidation or the recombination of related or heterologous genomes and of the proteins or peptides produced as a result thereof; thus, these genomes associated with the expression of  
5 and/or infection by MSRV-1 are an integral part of the potential pathogenicity of this virus, and hence constitute means of diagnostic detection and special therapeutic targets. Similarly, any agent associated with or cofactor of these interactions responsible for the  
10 pathogenesis in question, such as MSRV-2 or the gliotoxic factor which are described in the patent application published under No. FR-2,716,198, may participate in the development of an overall and very effective strategy for the diagnosis, prognosis, therapeutic monitoring and/or  
15 integrated therapy of MS in particular, but also of any other disease associated with the same agents.

In this context, a parallel discovery has been made in another autoimmune disease, rheumatoid arthritis (RA), which has been described in the French Patent  
20 Application filed under No. 95/02960. This discovery shows that, by applying methodological approaches similar to the ones which were used in the Applicant's work on MS, it was possible to identify a retrovirus expressed in RA which shares the sequences described for MSRV-1 in MS, and also  
25 the coexistence of an associated MSRV-2 sequence also described in MS. As regards MSRV-1, the sequences detected in common in MS and RA relate to the pol and gag genes. In the current state of knowledge, it is possible to associate the gag and pol sequences described with the  
30 MSRV-1 strains expressed in these two diseases.

The present patent application relates to various results which are additional to those already protected by the following French Patent Applications:

- No. 92/04322 of 03.04.1992, published under  
35 No. 2,689,519;

- No. 92/13447 of 03.11.1992, published under No. 2,689,521;
- No. 92/13443 of 03.11.1992, published under No. 2,689,520;
- 5 - No. 94/01529 of 04.02.1994, published under No. 2,715,936;
- No. 94/01531 of 04.02.1994, published under No. 2,715,939;
- No. 94/01530 of 04.02.1994, published under
- 10 No. 2,715,936;
- No. 94/01532 of 04.02.1994, published under No. 2,715,937;
- No. 94/14322 of 24.11.1994, published under No. 2,727,428;
- 15 - and No. 94/15810 of 23.12.1994; published under No. 2,728,585.

The present invention relates, in the first place, to a viral material, in the isolated or purified state, which may be recognized or characterized in

20 different ways:

- its genome comprises a nucleotide sequence chosen from the group including the sequences SEQ ID NO:46, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:56, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID
- 25 NO:89, their complementary sequences and their equivalent sequences, in particular nucleotide sequences displaying, for any succession of 100 contiguous monomers, at least 50% and preferably at least 70% homology with the said sequences SEQ ID NO:46, SEQ ID NO:51, SEQ ID NO:52, SEQ ID
- 30 NO:53, SEQ ID NO:56, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60 SEQ ID NO:61, SEQ ID NO:89, respectively, and their complementary sequences;
- the region of its genome comprising the env and pol genes and a portion of the gag gene, excluding the
- 35 subregion having a sequence identical or equivalent to SEQ ID NO:1, codes for any polypeptide displaying, for any

contiguous succession of at least 30 amino acids, at least 50% and preferably at least 70% homology with a peptide sequence encoded by any nucleotide sequence chosen from the group including SEQ ID NO:46, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:56, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60 SEQ ID NO:61 SEQ ID NO:89 and their complementary sequences;

- the pol gene comprises a nucleotide sequence partially or totally identical or equivalent to SEQ ID NO:57 or SEQ ID NO:93, excluding SEQ ID NO:1.

- the gag gene comprises a nucleotide sequence partially or totally identical or equivalent to SEQ ID NO:88.

As indicated above, according to the present invention, the viral material as defined above is associated with MS. And as defined by reference to the pol or gag gene of MSRV-1, and more especially to the sequences SEQ ID NOS 51, 56, 57, 59, 60, 61, 88, 89, 93, 169, 170, 171, 172, 176, 177, 178 and 179, this viral material is associated with RA.

The present invention also relates to a nucleic material, in the isolated or purified state, having at least one of the following definitions :

- a nucleic material comprising a nucleotide sequence selected from the group including sequences SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:169, SEQ ID NO:170, SEQ ID NO:171, SEQ ID NO:172, SEQ ID NO:176, SEQ ID NO:177, SEQ ID NO:178, SEQ ID NO:179, their complementary sequences and their equivalent sequences, in particular nucleotide sequences displaying, for any succession of 100 contiguous monomers, at least 50% and preferably at least 60% homology with said sequences SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:169, SEQ ID NO:170, SEQ ID NO:171, SEQ ID NO:172, SEQ ID NO:176, SEQ ID NO:177, SEQ ID NO:178, SEQ ID NO:179, and their complementary sequences, excluding HSERV-9 (or ERV-9) ; advantageously, the nucleotide sequence of said nucleic material is

selected from the group including sequences SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:169, SEQ ID NO:170, SEQ ID NO:171, SEQ ID NO:172, SEQ ID NO:176, SEQ ID NO:177, SEQ ID NO:178, SEQ ID NO:179, their complementary  
5 sequences and their equivalent sequences, in particular nucleotide sequences displaying, for any succession of 100 contiguous monomers, at least 70% and preferably at least 80% homology with said sequences SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:169, SEQ ID NO:170, SEQ ID NO:171,  
10 SEQ ID NO:172, SEQ ID NO:176, SEQ ID NO:177, SEQ ID NO:178, SEQ ID NO:179, and their complementary sequences ;

- a nucleic material, in the isolated or purified state, coding for any polypeptide displaying, for any contiguous  
15 succession of at least 30 amino acids, at least 50%, preferably at least 60 %, and most preferably at least 70% homology with a peptide sequence encoded by any nucleotide sequence selected from the group including SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:169, SEQ ID NO:170, SEQ ID NO:171,  
20 SEQ ID NO:172, SEQ ID NO:176, SEQ ID NO:177, SEQ ID NO:178, SEQ ID NO:179 and their complementary sequences;

- a nucleic material, in the isolated or purified state, of retroviral type, comprising a nucleotide sequence  
25 identical or similar to at least part of the pol gene of an isolated retrovirus associated with multiple sclerosis or rheumatoid arthritis; advantageously, said nucleotide sequence is 80 % similar to said at least part of the gene pol;

30 - a nucleic material comprising a nucleotide sequence identical or similar to at least part of the pol gen of an isolated virus encoding a reverse transcriptase having a enzymatic site comprised between the amino acid domains LPQG-YXDD, having a phylogenic distance with HSERV-9 of  
35  $0.063 \pm 0.1$ , and preferably  $0.063 \pm 0.05$ ; the phylogenic distances are calculated on the basis of a reference

sequence according to UPGM tree option of the Geneworks™ Software (INTELLIGENETICS) ;

By enzymatic site, we understand the amino acids domain(s) conferring the specific activity of a given enzyme.

5           The present invention also relates to different nucleotide fragments each comprising a nucleotide sequence chosen from the group including:

(a) all the genomic sequences, partial and total, of the pol gene of the MSRV-1 virus, except for the total  
10 sequence of the nucleotide fragment defined by SEQ ID NO:1;

(b) all the genomic sequences, partial and total, of the env gene of MSRV-1;

(c) all the partial genomic sequences of the gag gene of  
15 MSRV-1;

(d) all the genomic sequences overlapping the pol gene and the env gene of the MSRV-1 virus, and overlapping the pol gene and the gag gene;

(e) all the sequences, partial and total, of a clone  
20 chosen from the group including the clones FBd3 (SEQ ID NO:46), t pol (SEQ ID NO:51), JLBc1 (SEQ ID NO:52), JLBc2 (SEQ ID NO:53) and GM3 (SEQ ID NO:56), FBd13 (SEQ ID NO:58), LB19 (SEQ ID NO:59), LTRGAG12 (SEQ ID NO:60), FP6 (SEQ ID NO:61), G+E+A  
25 (SEQ ID NO:89), excluding any nucleotide sequence identical to or lying within the sequence defined by SEQ ID NO:1;

(f) sequences complementary to the said genomic sequences;

(g) sequences equivalent to the said sequences (a) to (e),  
30 in particular nucleotide sequences displaying, for any succession of 100 contiguous monomers, at least 50% and preferably at least 70% homology with the said sequences (a) to (d),

provided that this nucleotide fragment does not comprise  
35 or consist of the sequence ERV-9 as described in LA MANTIA et al. (18).



The term genomic sequences, partial or total, includes all sequences associated by coencapsidation or by coexpression, or recombined sequences.

Preferably, such a fragment comprises:

- 5 - either a nucleotide sequence identical to a partial or total genomic sequence of the pol gene of the MSRV-1 virus, except for the total sequence of the nucleotide fragment defined by SEQ ID NO:1, or identical to any sequence equivalent to the said partial or total genomic  
10 sequence, in particular one which is homologous to the latter;
  - or a nucleotide sequence identical to a partial or total genomic sequence of the env gene of the MSRV-1 virus, or identical to any sequence complementary to the said  
15 nucleotide sequence, or identical to any sequence equivalent to the said nucleotide sequence, in particular one which is homologous to the latter.

In particular, the invention relates to a nucleotide fragment comprising a coding nucleotide  
20 sequence which is partially or totally identical to a nucleotide sequence chosen from the group including:

- the nucleotide sequence defined by SEQ ID NO:40, SEQ ID NO:62 or SEQ ID NO:89;
- sequences complementary to SEQ ID NO:40, SEQ ID NO:62 or  
25 SEQ ID NO:89;
- sequences equivalent, and in particular homologous to SEQ ID NO:40, SEQ ID NO:62 or SEQ ID NO:89;
- sequences coding for all or part of the peptide sequence defined by SEQ ID NO:39, SEQ ID NO:63 or SEQ ID NO:90;
- 30 - sequences coding for all or part of a peptide sequence equivalent, in particular homologous to SEQ ID NO:39, SEQ ID NO:63 or SEQ ID NO:90, which is capable of being recognized by sera of patients infected with the MSRV-1 virus, or in whom the MSRV-1 virus has been reactivated.

The invention also relates to a nucleotide fragment (called fragment I) having at least one of the following definitions :

- a nucleotide fragment comprising a nucleotide sequence  
5 selected from the group including SEQ ID NO:93,  
SEQ ID NO:94, SEQ ID NO:169, SEQ ID NO:170, SEQ ID NO:171,  
SEQ ID NO:172, SEQ ID NO:176, SEQ ID NO:177,  
SEQ ID NO:178, SEQ ID NO:179, their complementary  
sequences and their equivalent sequences, in particular  
10 nucleotide sequences displaying, for any succession of 100  
contiguous monomers, at least 50% and preferably at least  
60% homology with said sequences and their complementary  
sequences, said group excluding SEQ ID NO:1,  
said nucleotide fragment not comprising nor consisting of  
15 the sequence HSERV-9 (or ERV-9); preferably the nucleotide  
sequence of said fragment is selected from the group  
including SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:169,  
SEQ ID NO:170, SEQ ID NO:171, SEQ ID NO:172,  
SEQ ID NO:176, SEQ ID NO:177, SEQ ID NO:178,  
20 SEQ ID NO:179, their complementary sequences and their  
equivalent sequences, in particular nucleotide sequences  
displaying, for any succession of 100 contiguous monomers,  
at least 70% and preferably at least 80% homology with  
said sequences and their complementary sequences;
- 25 - a nucleotide fragment comprising a coding nucleotide  
sequence which is partially or totally identical to a  
nucleotide sequence selected from the group including :  
SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:169,  
SEQ ID NO:170, SEQ ID NO:171, SEQ ID NO:172,  
30 SEQ ID NO:176, SEQ ID NO:177, SEQ ID NO:178,  
SEQ ID NO:179 ; their complementary sequences ; their  
equivalent sequences, in particular homologous to  
SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:169, SEQ ID NO:170,  
SEQ ID NO:171, SEQ ID NO:172, SEQ ID NO:176,  
35 SEQ ID NO:177, SEQ ID NO:178, SEQ ID NO:179;

sequences encoding all or parts of the peptide sequence defined by SEQ ID NO:95, SEQ ID NO:173, SEQ ID NO:174, SEQ ID NO:175, SEQ ID NO:180, SEQ ID NO:181, SEQ ID NO:182;

5 sequences encoding all or parts of a peptide sequence equivalent, in particular homologous to SEQ ID NO:95, SEQ ID NO:173, SEQ ID NO:174, SEQ ID NO:175, SEQ ID NO:180, SEQ ID NO:181, SEQ ID NO:182, which is capable of being recognized by sera of patients infected  
10 with the MSRV-1 virus, or in whom the MSRV-1 virus has been reactivated.

The invention also relates to any nucleic acid probe for the detection of virus associated with MS and/or rheumatoid arthritis (RA), which is capable of hybridizing  
15 specifically with any fragment such as is defined above, belonging or lying within the genome of the said pathogenic agent. It relates, in addition, to any nucleic acid probe for detection of a pathogenic and/or infective agent associated with RA, which is capable of hybridizing  
20 specifically with any fragment as defined above by reference to the pol and gag genes, and especially with respect to the sequences SEQ ID NOS 40, 51, 56, 59, 60, 61, 62, 89 and SEQ ID NOS 39, 63 and 90.

The invention also relates to a primer for the  
25 amplification by polymerization of an RNA or a DNA of a viral material, associated with MS and/or RA, comprising a nucleotide sequence identical or equivalent to at least one portion of the nucleotide sequence of any fragment such as is defined above, in particular a nucleotide  
30 sequence displaying, for any succession of at least 10 contiguous monomers, preferably 15 contiguous monomers, more preferably 18 contiguous monomers and even most preferably 20 contiguous monomers, at least 70% homology with at least the said portion of the said fragment.  
35 Preferably, the nucleotide sequence of such a primer is identical to any one of the sequences selected from the

group including SEQ ID NO:47 to SEQ ID NO:50, SEQ ID NO:55, SEQ ID NO:64, SEQ ID NO:86, SEQ ID NO:99 to SEQ ID NO:111, SEQ ID NO:183, SEQ ID NO:184, SEQ ID NO:185, SEQ ID NO:186.

5 Generally speaking the invention also encompasses any RNA or DNA, and in particular replication vector, comprising a genomic fragment of the viral material such as is defined above, or a nucleotide fragment such as is defined above.

10 The invention also relates to the different peptides encoded by any open reading frame belonging to a nucleotide fragment such as is defined above, in particular any polypeptide, for example any oligopeptide forming or comprising an antigenic determinant recognized  
15 by sera of patients infected with the MSRV-1 virus and/or in whom the MSRV-1 virus has been reactivated. Preferably, this polypeptide is antigenic, and is encoded by the open reading frame beginning, in the 5'-3' direction, at nucleotide 181 and ending at nucleotide 330 of  
20 SEQ ID NO:1.

The invention also encompasses the following polypeptides :

a)

- a polypeptide encoded by any open reading frame  
25 belonging to a nucleotide fragment, fragment I, as defined above ;

- a polypeptide, characterized in that the open reading frame encoding it, is comprised, in the 5'-3' direction, between nucleotide 18 and nucleotide 2304 of SEQ ID NO:93;

30 - a polypeptide, having a peptide sequence comprising a sequence partially or totally identical to SEQ ID NO:95;

b)

- a polypeptide, recombinant or synthetic, having a peptide sequence which comprises a sequence identical or  
35 equivalent to SEQ ID NO:96; in particular said polypeptide

exhibits an enzymatic activity consisting of proteolytic activity;

- a polypeptide, recombinant or synthetic, characterized in that the open reading frame encoding it begins, in the  
5 5'-3' direction, at nucleotide 18 and ends at nucleotide 340 of SEQ ID NO:93;

- a polypeptide having an inhibitory activity on the proteolytic activity of a polypeptide as defined according to b);

10 c)

- a polypeptide, recombinant or synthetic, having a peptide sequence which comprises a sequence identical or equivalent to SEQ ID NO:97; in particular said polypeptide exhibits a reverse transcriptase activity;

15 - a polypeptide having a peptide sequence which comprises a sequence identical or equivalent to SEQ ID NO:98; in particular said polypeptide exhibits a ribonuclease activity;

- a polypeptide, recombinant or synthetic, characterized  
20 in that the open reading frame encoding it begins, in the 5'-3' direction, at nucleotide 341 and ends at nucleotide 2304 of SEQ ID NO:93;

- a polypeptide, recombinant or synthetic, characterized in that the open reading frame encoding it begins, in the  
25 5'-3' direction, at nucleotide 1858 and ends at nucleotide 2304 of SEQ ID NO:93.

- a polypeptide having an inhibitory activity on the reverse transcriptase activity of a polypeptide as defined according to c) or on the ribonuclease H activity of a  
30 polypeptide as defined according to c).

In particular, the invention relates to an antigenic polypeptide recognized by the sera of patients infected with the MSRV-1 virus, and/or in whom the MSRV-1 virus has been reactivated, whose peptide sequence is  
35 partially or totally identical or is equivalent to the sequence defined by SEQ ID NO:39, SEQ ID NO:63,

SEQ ID NO:87, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:173, SEQ ID NO:174, SEQ ID NO:175, SEQ ID NO:180, SEQ ID NO:181 and SEQ ID NO:182; such a sequence is identical, for example, to any sequence  
5 selected from the group including the sequences SEQ ID NO:41 to SEQ ID NO:44, SEQ ID NO:63 and SEQ ID NO:87.

The present invention also proposes mono- or polyclonal antibodies directed against the MSRV-1 virus,  
10 which are obtained by the immunological reaction of a human or animal body or cells to an immunogenic agent consisting of an antigenic polypeptide such as is defined above.

The invention next relates to:

- 15 - reagents for detection of the MSRV- virus, or of an exposure to the latter, comprising, at least one reactive substance selected from the group consisting of a probe of the present invention, a polypeptide, in particular an antigenic peptide, such as is defined above, or an anti-  
20 ligand, in particular an antibody to the said polypeptide;  
- all diagnostic, prophylactic or therapeutic compositions comprising one or more peptides, in particular antigenic peptides, such as are defined above, or one or more anti-ligands, in particular antibodies to the peptides,  
25 discussed above; such a composition is preferably, and by way of example, a vaccine composition.

The invention also relates to any diagnostic, prophylactic or therapeutic composition, in particular for inhibiting the expression of at least one virus associated  
30 with MS or RA, and/or the enzymatic activities of the proteins of said virus, comprising a nucleotide fragment such as is defined above or a polynucleotide, in particular oligonucleotide, whose sequence is partially identical to that of the said fragment, except for that of  
35 the fragment having the nucleotide sequence SEQ ID NO:1. Likewise, it relates to any diagnostic, prophylactic or

therapeutic composition, in particular for inhibiting the expression of at least one pathogenic and/or infective agent associated with RA, comprising a nucleotide fragment such as is defined above by reference to the pol and gag  
5 genes, and especially with respect to the sequences SEQ ID NOS 40, 51, 56, 59, 60, 61, 62 and 89.

According to the invention, these same fragments or polynucleotides, in particular oligonucleotides, may participate in all suitable compositions for detecting,  
10 according to any suitable process or method, a pathological and/or infective agent associated with MS and with RA, respectively, in a biological sample. In such a process, an RNA and/or a DNA presumed to belong or originating from the said pathological and/or infective  
15 agent, and/or their complementary RNA and/or DNA, is/are brought into contact with such a composition.

The present invention also relates to any process for detecting the presence or exposure to such a pathological and/or infective agent, in a biological  
20 sample, by bringing this sample into contact with a peptide, in particular an antigenic peptide such as is defined above, or an anti-ligand, in particular an antibody to this peptide, such as is defined above.

In practice, and for example, a device for  
25 detection of the MSRV-1 virus comprises a reagent such as is defined above, supported by a solid support which is immunologically compatible with the reagent, and a means for bringing the biological sample, for example a sample of blood or of cerebrospinal fluid, likely to contain  
30 anti-MSRV-1 antibodies, into contact with this reagent under conditions permitting a possible immunological reaction, the foregoing items being accompanied by means for detecting the immune complex formed with this reagent.

Lastly, the invention also relates to the detec-  
35 tion of anti-MSRV-1 antibodies in a biological sample, for example a sample of blood or of cerebrospinal fluid,

according to which this sample is brought into contact with a reagent such as is defined above, consisting of an antibody, under conditions permitting their possible immunological reaction, and the presence of the immune  
5 complex thereby formed with the reagent is then detected.

Before describing the invention in detail, different terms used in the description and the claims are now defined:

- strain or isolate is understood to mean any  
10 infective and/or pathogenic biological fraction containing, for example, viruses and/or bacteria and/or parasites, generating pathogenic and/or antigenic power, harboured by a culture or a living host; as an example, a viral strain according to the above definition can contain  
15 a coinfective agent, for example a pathogenic protist,

- the term "MSRV" used in the present description denotes any pathogenic and/or infective agent associated with MS, in particular a viral species, the attenuated strains of the said viral species or the  
20 defective-interfering particles or particles containing coencapsidated genomes, or alternatively genomes recombined with a portion of the MSRV-1 genome, derived from this species. Viruses, and especially viruses containing RNA, are known to have a variability resulting,  
25 in particular, from relatively high rates of spontaneous mutation (7), which will be borne in mind below for defining the notion of equivalence,

- human virus is understood to mean a virus capable of infecting, or of being harboured by human  
30 beings,

- in view of all the natural or induced variations and/or recombination which may be encountered when implementing the present invention, the subjects of the latter, defined above and in the claims, have been  
35 expressed including the equivalents or derivatives of the different biological materials defined below, in



particular of the homologous nucleotide or peptide sequences,

- the variant of a virus or of a pathogenic and/or infective agent according to the invention  
5 comprises at least one antigen recognized by at least one antibody directed against at least one corresponding antigen of the said virus and/or said pathogenic and/or infective agent, and/or a genome any part of which is detected by at least one hybridization probe and/or at  
10 least one nucleotide amplification primer specific for the said virus and/or pathogenic and/or infective agent, such as, for example, for the MSRV-1 virus, the primers and probes having a nucleotide sequence chosen from SEQ ID NO:20 to SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:16  
15 to SEQ ID NO:19, SEQ ID NO:31 to SEQ ID NO:33, SEQ ID NO:45, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:45 and their complementary sequences, under particular hybridization conditions well known to a person skilled in the art,

20 - according to the invention, a nucleotide fragment or an oligonucleotide or polynucleotide is an arrangement of monomers, or a biopolymer, characterized by the informational sequence of the natural nucleic acids, which is capable of hybridizing with any other nucleotide  
25 fragment under predetermined conditions, it being possible for the arrangement to contain monomers of different chemical structures and to be obtained from a molecule of natural nucleic acid and/or by genetic recombination and/or by chemical synthesis; a nucleotide fragment may be  
30 identical to a genomic fragment of the MSRV-1 virus discussed in the present invention, in particular a gene of this virus, for example pol or env in the case of the said virus,

- thus, a monomer can be a natural nucleotide of  
35 nucleic acid whose constituent elements are a sugar, a phosphate group and a nitrogenous base; in RNA the sugar

is ribose, in DNA the sugar is 2-deoxyribose; depending on whether the nucleic acid is DNA or RNA, the nitrogenous base is chosen from adenine, guanine, uracil, cytosine and thymine; or the nucleotide can be modified in at least one of the three constituent elements; as an example, the modification can occur in the bases, generating modified bases such as inosine, 5-methyldeoxycytidine, deoxyuridine, 5-(dimethylamino)deoxyuridine, 2,6-diaminopurine, 5-bromodeoxyuridine and any other modified base promoting hybridization; in the sugar, the modification can consist of the replacement of at least one deoxyribose by a polyamide (8), and in the phosphate group, the modification can consist of its replacement by esters chosen, in particular, from diphosphate, alkyl- and arylphosphonate and phosphorothioate esters,

- "informational sequence" is understood to mean any ordered succession of monomers whose chemical nature and order in a reference direction constitute or otherwise an item of functional information of the same quality as that of the natural nucleic acids,

- hybridization is understood to mean the process during which, under suitable working conditions, two nucleotide fragments having sufficiently complementary sequences pair to form a complex structure, in particular double or triple, preferably in the form of a helix,

- a probe comprises a nucleotide fragment synthesized chemically or obtained by digestion or enzymatic cleavage of a longer nucleotide fragment, comprising at least six monomers, advantageously from 10 to 1000 monomers, preferably 10 to 30 monomers and more preferably 18 to 30, and possessing a specificity of hybridization under particular conditions; preferably, a probe possessing fewer than 10 monomers, but preferably fewer than 15 monomers is not used alone, but is used in the presence of other probes of equally short size or otherwise; under certain special conditions, it may be useful to use probes

of size greater than 100 monomers; a probe may be used, in particular, for diagnostic purposes, such molecules being, for example, capture and/or detection probes,

- the capture probe may be immobilized on a solid support by any suitable means, that is to say directly or indirectly, for example by covalent bonding or passive adsorption,

- the detection probe may be labelled by means of a label chosen, in particular, from radioactive isotopes, enzymes chosen, in particular, from peroxidase and alkaline phosphatase and those capable of hydrolysing a chromogenic, fluorogenic or luminescent substrate, chromophoric chemical compounds, chromogenic, fluorogenic or luminescent compounds, nucleotide base analogues and biotin,

- the probes used for diagnostic purposes of the invention may be employed in all known hybridization techniques, and in particular the techniques termed "DOT-BLOT" (9), "SOUTHERN BLOT" (10), "NORTHERN BLOT", which is a technique identical to the "SOUTHERN BLOT" technique but which uses RNA as target, and the SANDWICH technique (11); advantageously, the SANDWICH technique is used in the present invention, comprising a specific capture probe and/or a specific detection probe, on the understanding that the capture probe and the detection probe must possess an at least partially different nucleotide sequence,

- any probe according to the present invention can hybridize in vivo or in vitro with RNA and/or with DNA in order to block the phenomena of replication, in particular translation and/or transcription, and/or to degrade the said DNA and/or RNA,

- a primer is a probe comprising at least six monomers, and advantageously from 10 to 30 monomers, and preferably from 18 to 25 monomers, possessing a specificity of hybridization under particular conditions

for the initiation of an enzymatic polymerization, for example in an amplification technique such as PCR (polymerase chain reaction), in an elongation process such as sequencing, in a method of reverse transcription or the like,

- two nucleotide or peptide sequences are termed equivalent or derived with respect to one another, or with respect to a reference sequence, if functionally the corresponding biopolymers can perform substantially the same role, without being identical, as regards the application or use in question, or in the technique in which they participate; two sequences are, in particular, equivalent if they are obtained as a result of natural variability, in particular spontaneous mutation of the species from which they have been identified, or induced variability, as are two homologous sequences, homology being defined below,

- "variability" is understood to mean any spontaneous or induced modification of a sequence, in particular by substitution and/or insertion and/or deletion of nucleotides and/or of nucleotide fragments, and/or extension and/or shortening of the sequence at one or both ends; an unnatural variability can result from the genetic engineering techniques used, for example the choice of synthesis primers, degenerate or otherwise, selected for amplifying a nucleic acid; this variability can manifest itself in modifications of any starting sequence, considered as reference, and capable of being expressed by a degree of homology relative to the said reference sequence,

- homology characterizes the degree of identity of two nucleotide or peptide fragments compared; it is measured by the percentage identity which is determined, in particular, by direct comparison of nucleotide or peptide sequences, relative to reference nucleotide or peptide sequences,

- this percentage identity has been specifically determined for the nucleotide fragments, clones in particular, dealt with in the present invention, which are homologous to the fragments identified, for the MSRV-1 virus, by SEQ ID NO:1 to NO:9, SEQ ID NO:46, SEQ ID NO:51 to SEQ ID NO:53, SEQ ID NO:40, SEQ ID NO:56, SEQ ID NO:57 and SEQ ID NO:93, as well as for the probes and primers homologous to the probes and primers identified by SEQ ID NO:20 to SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:16 to SEQ ID NO:19, SEQ ID NO:31 to SEQ ID NO:33, SEQ ID NO:45, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:55, SEQ ID NO:40, SEQ ID NO:56, SEQ ID NO:57 and SEQ ID NO:99 to SEQ ID NO:111; as an example, the smallest percentage identity observed between the different general consensus sequences of nucleic acids obtained from fragments of MSRV-1 viral RNA, originating from the LM7PC and PLI-2 lines according to a protocol detailed later, is 67% in the region described in Figure 1,

- any nucleotide fragment is termed equivalent or derived from a reference fragment if it possesses a nucleotide sequence equivalent to the sequence of the reference fragment; according to the above definition, the following in particular are equivalent to a reference nucleotide fragment:

a) any fragment capable of hybridizing at least partially with the complement of the reference fragment,

b) any fragment whose alignment with the reference fragment results in the demonstration of a larger number of identical contiguous bases than with any other fragment originating from another taxonomic group,

c) any fragment resulting, or capable of resulting, from the natural variability of the species from which it is obtained,

d) any fragment capable of resulting from the genetic engineering techniques applied to the reference fragment,

e) any fragment containing at least eight contiguous nucleotides encoding a peptide which is homologous or identical to the peptide encoded by the reference fragment,

5 f) any fragment which is different from the reference fragment by insertion, deletion or substitution of at least one monomer, or extension or shortening at one or both of its ends; for example, any fragment corresponding to the reference fragment flanked at one or  
10 both of its ends by a nucleotide sequence not coding for a polypeptide,

- polypeptide is understood to mean, in particular, any peptide of at least two amino acids, in particular an oligopeptide, or protein, and for example an  
15 enzyme, extracted, separated or substantially isolated or synthesized through human intervention, in particular those obtained by chemical synthesis or by expression in a recombinant organism,

- polypeptide partially encoded by a nucleotide  
20 fragment is understood to mean a polypeptide possessing at least three amino acids encoded by at least nine contiguous monomers lying within the said nucleotide fragment,

- an amino acid is termed analogous to another  
25 amino acid when their respective physicochemical properties, such as polarity, hydrophobicity and/or basicity and/or acidity and/or neutrality are substantially the same; thus, a leucine is analogous to an isoleucine.

- any polypeptide is termed equivalent or  
30 derived from a reference polypeptide if the polypeptides compared have substantially the same properties, and in particular the same antigenic, immunological, enzymological and/or molecular recognition properties; the following in particular are equivalent to a reference  
35 polypeptide:

a) any polypeptide possessing a sequence in which at least one amino acid has been replaced by an analogous amino acid,

b) any polypeptide having an equivalent peptide sequence, obtained by natural or induced variation of the said reference polypeptide and/or of the nucleotide fragment coding for the said polypeptide,

c) a mimotope of the said reference polypeptide,

d) any polypeptide in whose sequence one or more amino acids of the L series are replaced by an amino acid of the D series, and vice versa,

e) any polypeptide into whose sequence a modification of the side chains of the amino acids has been introduced, such as, for example, an acetylation of the amine functions, a carboxylation of the thiol functions, an esterification of the carboxyl functions,

f) any polypeptide in whose sequence one or more peptide bonds have been modified, such as, for example, carba, retro, inverso, retro-inverso, reduced and methylenoxy bonds,

(g) any polypeptide at least one antigen of which is recognized by an antibody directed against a reference polypeptide,

- the percentage identity characterizing the homology of two peptide fragments compared is, according to the present invention, at least 50% and preferably at least 70%.

In view of the fact that a virus possessing reverse transcriptase enzymatic activity may be genetically characterized equally well in RNA and in DNA form, both the viral DNA and RNA will be referred to for characterizing the sequences relating to a virus possessing such reverse transcriptase activity, termed MSRV-1 according to the present description.

The expressions of order used in the present description and the claims, such as "first nucleotide

sequence", are not adopted so as to express a particular order, but so as to define the invention more clearly.

Detection of a substance or agent is understood below to mean both an identification and a quantification, or a separation or isolation, of the said substance or said agent.

A better understanding of the invention will be gained on reading the detailed description which follows, prepared with reference to the attached figures, in which:

10       - Figure 1 shows general consensus sequences of nucleic acids of the MSRV-1B clones amplified by the PCR technique in the "pol" region defined by Shih (12), from viral DNA originating from the LM7PC and PLI-2 lines, and identified under the references SEQ ID NO:3, SEQ ID NO:4, 15 SEQ ID NO:5 and SEQ ID NO:6, and the common consensus with amplification primers bearing the reference SEQ ID NO:7;

      - Figure 2 gives the definition of a functional reading frame for each MSRV-1B/"PCR pol" type family, the said families A to D being defined, respectively, by the 20 nucleotide sequences SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5 and SEQ ID NO:6 described in Figure 1;

      - Figure 3 gives an example of consensus of the MSRV-2B sequences, identified by SEQ ID NO:11;

25       - Figure 4 is a representation of the reverse transcriptase (RT) activity in dpm (disintegrations per minute) in the sucrose fractions taken from a purification gradient of the virions produced by the B lymphocytes in culture from a patient suffering from MS;

30       - Figure 5 gives, under the same experimental conditions as in Figure 4, the assay of the reverse transcriptase activity in the culture of a B lymphocyte line obtained from a control free from MS;

      - Figure 6 shows the nucleotide sequence of the clone PSJ17 (SEQ ID NO:9);

35       - Figure 7 shows the nucleotide sequence SEQ ID NO:8 of the clone designated M003-P004;



- Figure 8 shows the nucleotide sequence SEQ ID NO:2 of the clone F11-1; the portion located between the two arrows in the region of the primer corresponds to a variability imposed by the choice of primer which was used for the cloning of F11-1; in this same figure, the translation into amino acids is shown;

- Figure 9 shows the nucleotide sequence SEQ ID NO:1, and a possible functional reading frame of SEQ ID NO:1 in terms of amino acids; on this sequence, the consensus sequences of the pol gene are underlined;

- Figures 10 and 11 give the results of a PCR, in the form of a photograph under ultraviolet light of an ethidium bromide-impregnated agarose gel, of the amplification products obtained from the primers identified by SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18 and SEQ ID NO:19;

- Figure 12 gives a representation in matrix form of the homology between SEQ ID NO:1 of MSRV-1 and that of an endogenous retrovirus designated HSERV9; this homology of at least 65% is demonstrated by a continuous line, the absence of a line meaning a homology of less than 65%;

- Figure 13 shows the nucleotide sequence SEQ ID NO:46 of the clone FBd3;

- Figure 14 shows the sequence homology between the clone FBd3 and the HSERV-9 retrovirus;

- Figure 15 shows the nucleotide sequence SEQ ID NO:51 of the clone t pol;

- Figures 16 and 17 show, respectively, the nucleotide sequences SEQ ID NO:52 and SEQ ID NO:53 of the clones JLBc1 and JLBc2, respectively;

- Figure 18 shows the sequence homology between the clone JLBc1 and the clone FBd3;

- and Figure 19 the sequence homology between the clone JLBc2 and the clone FBd3;

- Figure 20 shows the sequence homology between the clones JLBc1 and JLBc2;

- Figures 21 and 22 show the sequence homology between the HSERV-9 retrovirus and the clones JLBc1 and JLBc2, respectively;

- Figure 23 shows the nucleotide sequence SEQ ID  
5 NO:56 of the clone GM3;

- Figure 24 shows the sequence homology between the HSERV-9 retrovirus and the clone GM3;

- Figure 25 shows the localization of the different clones studied, relative to the genome of the  
10 known retrovirus ERV9;

- Figure 26 shows the position of the clones F11-1, M003-P004, MSRV-1B and PSJ17 in the region hereinafter designated MSRV-1 pol\*;

- Figure 27, split into three successive Figures  
15 27a-27c, shows a possible reading frame covering the whole of the pol gene;

- Figure 28 shows, according to SEQ ID NO:40, the nucleotide sequence coding for the peptide fragment POL2B, having the amino acid sequence identified by SEQ ID  
20 NO:39;

- Figure 29 shows the OD values (ELISA tests) at 492 nm obtained for 29 sera of MS patients and 32 sera of healthy controls tested with an anti-IgG antibody;

- Figure 30 shows the OD values (ELISA tests) at  
25 492 nm obtained for 36 sera of MS patients and 42 sera of healthy controls tested with an anti-IgM antibody;

- Figures 31 to 33 show the results obtained (relative intensity of the spots) for 43 overlapping octapeptides covering the amino acid sequence 61-110,  
30 according to the Spotscan technique, respectively with a pool of MS sera, with a pool of control sera and with the pool of MS sera after deduction of a background corresponding to the maximum signal detected on at least one octapeptide with the control serum (intensity = 1), on the  
35 understanding that these sera were diluted to 1/50. The

bar at the far right-hand end represents a graphic scale standard unrelated to the serological test;

- Figure 34 shows the SEQ ID NO:41 and SEQ ID NO:42 of two polypeptides comprising immunodominant regions, while SEQ ID NO:43 and 44 represent immunoreactive polypeptides specific to MS;

- Figure 35 shows the nucleotide sequence SEQ ID NO:59 of the clone LB19 and three potential reading frames of SEQ ID NO:59 in terms of amino acids;

10 - Figure 36 shows the nucleotide sequence SEQ ID NO:88 (GAG\*) and a potential reading frame of SEQ ID NO:88 in terms of amino acids;

- Figure 37 shows the sequence homology between the clone FBd13 and the HSERV-9 retrovirus; according to 15 this representation, the continuous line means a percentage homology greater than or equal to 70% and the absence of a line means a smaller percentage homology;

- Figure 38 shows the nucleotide sequence SEQ ID NO:61 of the clone FP6 and three potential reading frames 20 of SEQ ID NO:61 in terms of amino acids;

- Figure 39 shows the nucleotide sequence SEQ ID NO:89 of the clone G+E+A and three potential reading frames of SEQ ID NO:89 in terms of amino acids;

- Figure 40 shows a reading frame found in the 25 region E and coding for an MSRV-1 retroviral protease identified by SEQ ID NO:90;

- Figure 41 shows the response of each serum of patients suffering from MS, indicated by the symbol (+), and of healthy patients, symbolised by (-), tested with an anti-IgG antibody, expressed as net optical density at 30 492 nm;

- Figure 42 shows the response of each serum of patients suffering from MS, indicated by the symbols (+) and (QS), and of healthy patients (-), tested with an 35 anti-IgM antibody, expressed as net optical density at 492 nm;

- Figure 43 shows the RT-activity profile in sucrose density gradients of pellets from B-cell lines supernatants; Control B-cell line ■ was obtained from the relative of a patient with mitochondriopathy. MS B-Cell  
5 line □ was obtained from a patient with definite MS;

- Figure 44 shows the nucleotide and amino acid alignment of the conserved *pol* regions of viruses detected in the study (cf Example 18) by the "Pan-retrovirus" PCR. "Deletions" are represented by dashes and standard single-  
10 letter abbreviations are used to designate amino acids and nucleotides (i = inosine). The most highly conserved VLPQG and YXDD regions are shown as separate blocks in bold type at the end of each sequence. Amino acids which are present in all or in all but one of the sequences are underlined.  
15 PCR primers (modified from (12)) PAN-UO and PAN-UI are orientated 5' to 3' (sense) whereas primer PAN-DI is 3' to 5' (antisense). Degeneracies are shown above (PAN-UO & PAN-DI) or below (PAN-UI) the PCR primer sequences. "I" denotes the nine base 5' extension *cttggatcc*, "-I" denotes the nine base 5' extension *ctcaagctt*. The capture and detector probes DpV1 and CpV1b used in the ELOSA assay are shown below a representative MSRV-*cpol* sequence. At three positions below the translated MSRV-*cpol* sequence alternative amino acids (representing "non-silent" nucleic  
25 acid variations) are shown in italics - *K* and *Y* substitutions were only observed in PLI-1 derived clones whereas *R* and *W* were encoded by a significant proportion of the clones irrespective of derivation. Note that DpV1 is peroxidase labelled and that CpV1b may be biotinylated at the 5' end if streptavidin coated plates are used. The name of each sequence is indicated at the left of the figure.

HTLV1: Human Leukaemia Virus type 1; HIV1: Human Immunodeficiency Virus type 1; MoMLV: Moloney-Murine  
35 Leukaemia Virus; MPMV: Mason-Pfizer Monkey Virus. ERV9:

Endogenous Retrovirus 9. MSRV-cpol: Multiple Sclerosis associated RetroVirus conserved pol region.

- Figure 45 shows a phylogenic tree which is based on the conserved amino acid region encoded by the pol gene of MSRV and of representative endogenous and exogenous retroviruses and DNA viruses with reverse transcriptase. It was generated by the U.P.G.M.A. tree program of Geneworks® software.

HSRV: Human Spumaretrovirus. EIAV: Equine Infectious Aenemia Virus. BLV: Bovine Leukaemia Virus. HIV1, HIV2: Human Immunodeficiency Viruses type 1 and 2. HTLV1 and HTLV2: Human Leukaemia Viruses type 1 and 2. F-MuLV: Friend-Murine Leukaemia Virus. MoMLV: Moloney-Murine Leukaemia Virus. BAEV: Baboon Endogenous Virus. GaLV/ Gibbon Ape Leukaemia Virus. HUMER41: Human Endogenous Retroviral sequence, clone 41. IAP: Intracisternal A-type Particle. MPMV: Mason-Pfizer Monkey Virus. HERVK10: Human Endogenous Retrovirus K10. MMTV: Mouse Mammary tumour Virus. HSERV9 (ERV9 database sequence): Human sequence of Endogenous Retrovirus 9. MSRV: Multiple Sclerosis associated RetroVirus. SIV: Simian Immunodeficiency Virus; RTLH-H: Reverse Transcriptase-Like Viral sequence H; SFV: Simian Foamy Virus; VISNA: Visna retrovirus; SIV1: Simian Immunodeficiency Virus type 1; SRV-2: Simian Retrovirus type 2; SMRV-H: Squirrel Monkey Retrovirus H.

- Figure 46 shows the MSRV sequence in the Protease and Reverse-Transcriptase regions of the pol gene.

The aminoacid translation is aligned under the corresponding nucleotide sequence. The region corresponding to the Protease ORF cloned in a recombinant vector and expressed in *E. coli*, is boxed. The regions corresponding to the A and B fragments amplified on plasma samples from MS patients are indicated by brackets. The Reverse-Transcriptase (RT) and RNase H (RNH) region is boxed with dotted line. The highly conserved aminoacids

and/or active sites of enzyme activities of both PRT and RT (including RNH) are shown underlined.

- Figure 47A illustrates the specific detection of MSRV-pol RNA sequence by RT-PCR in the sucrose density fraction associated with RT-activity and in MS plasma ; Figure 47B shows the RT-activity profile on a sucrose density gradient obtained with extracellular virion pelleted from an MS choroid-plexus culture. The photograph below shows an agarose gel loaded with PCR products amplified from round 1 (ST1.1) RT-PCR products with the ST1.2 primer set. From left to right: water control 1 from RT-PCR step with ST1.1 set; water control 2 amplified from water control 1 with ST1.2 nested primers; Molecular weight markers; Fraction n°1 to 10 corresponding to the RT-activity profile shown above; Plasma samples C1 and C2 from healthy blood donors. Plasma samples MS1 and MS2 from two MS patients.

- Figure 48 shows an example of a variant and/or recombined sequence in the region of the pol gene defined by homology with the overlapping regions described in Figure 25, as GM3, MSRV-1 pol\*, t pol and FBd3.

- Figure 49 shows the nucleotide (Figure 49A) and amino acid (Figure 49B) alignments of the pol region between clones 1, 5 and 8 of the same patient (Experiment 46-7).

- Figure 50 shows the nucleotide (Figure 50A) and amino acid (Figure 50B) alignments of the pol region between clones 41, 43 and 42 of the same patient (Experiment 68-1).

- Figure 51 shows the nucleotide (Figure 51A) and amino acid (Figure 51B) alignments of the pol region between the consensus sequence (SEQ ID NO: 176) of clones 1, 5 and 8 of the same patient (Experiment 46-7) and SEQ ID NO:1, and between their corresponding peptide sequences.

- Figure 52 shows the nucleotide (Figure 52A) and amino acid (Figure 52B) alignments of the *pol* region between the consensus sequence (SEQ ID NO: 169) of clones 41, 43 and 42 of the same patient (Experiment 68-1) and  
5 SEQ ID NO:1, and between their corresponding peptide sequences.

- Figure 53 shows the nucleotide (Figure 53A) and amino acid (Figure 53B) alignments of the *pol* region between the consensus sequence (SEQ ID NO: 176) of clones  
10 1, 5 and 8 of the same patient (Experiment 46-7) and the consensus sequence (SEQ ID NO: 169) of clones 41, 43 and 42 of the same patient (Experiment 68-1).

Table 5 (at the end of the description) shows the sequences obtained by RT-PCR with degenerate *pol*  
15 primers on sucrose density gradient fractions containing the peak of RT-activity or its negative control (cf Example 18) ; and

Table 6 (at the end of the description) shows the clinical data and results of MSRV-c<sub>pol</sub> detection by  
20 "Pan-retro" PCR with specific ELOSA assay, on CSF from MS and control patients (cf Example 18).

**EXAMPLE 1: OBTAINING CLONES DESIGNATED MSRV-1B AND MSRV-2B, DEFINING, RESPECTIVELY, A RETROVIRUS MSRV-1  
25 AND A COINFECTION AGENT MSRV2, BY "NESTED" PCR AMPLIFICATION OF THE CONSERVED POL REGIONS OF RETROVIRUSES ON VIRION PREPARATIONS ORIGINATING FROM THE LM7PC AND PLI-2 LINES**

A PCR technique derived from the technique  
30 published by Shih (12) was used. This technique enables all trace of contaminant DNA to be removed by treating all the components of the reaction medium with DNase. It concomitantly makes it possible, by the use of different but overlapping primers in two successive series of PCR  
35 amplification cycles, to increase the chances of amplifying a cDNA synthesized from an amount of RNA which is

small at the outset and further reduced in the sample by the spurious action of the DNase on the RNA. In effect, the DNase is used under conditions of activity in excess which enable all trace of contaminant DNA to be removed before inactivation of this enzyme remaining in the sample by heating to 85°C for 10 minutes. This variant of the PCR technique described by Shih (12) was used on a cDNA synthesized from the nucleic acids of fractions of infective particles purified on a sucrose gradient according to the technique described by H. Perron (13) from the "POL-2" isolate (ECACC No. V92072202) produced by the PLI-2 line (ECACC No. 92072201) on the one hand, and from the MS7PG isolate (ECACC No. V93010816) produced by the LM7PC line (ECACC No. 93010817) on the other hand. These cultures were obtained according to the methods which formed the subject of the patent applications published under Nos WO 93/20188 and WO 93/20189.

After cloning the products amplified by this technique with the TA Cloning Kit® and analysis of the sequence using an Applied Biosystems model 373A Automatic Sequencer, the sequences were analysed using the Geneworks® software on the latest available version of the Genebank® data bank.

The sequences cloned and sequenced from these samples correspond, in particular, to two types of sequence: a first type of sequence, to be found in the majority of the clones (55% of the clones originating from the POL-2 isolates of the PLI-2 culture, and 67% of the clones originating from the MS7PG isolates of the LM7PC cultures), which corresponds to a family of "pol" sequences closely similar to, but different from, the endogenous human retrovirus designated ERV-9 or HSERV-9, and a second type of sequence which corresponds to sequences very strongly homologous to a sequence attributed to another infective and/or pathogenic agent designated MSRV-2.



The first type of sequence, representing the majority of the clones, consists of sequences whose variability enables four subfamilies of sequences to be defined. These subfamilies are sufficiently similar to one another for it to be possible to consider them to be quasi-species originating from the same retrovirus, as is well known for the HIV-1 retrovirus (14), or to be the outcome of interference with several endogenous proviruses coregulated in the producing cells. These more or less defective endogenous elements are sensitive to the same regulatory signals possibly generated by a replicative provirus, since they belong to the same family of endogenous retroviruses (15). This new family of endogenous retroviruses, or alternatively this new retroviral species from which the generation of quasi-species has been obtained in culture, and which contains a consensus of the sequences described below, is designated MSRV-1B.

Figure 1 presents the general consensus sequences of the sequences of the different MSRV-1B clones sequenced in this experiment, these sequences being identified, respectively, by SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5 and SEQ ID NO:6. These sequences display a homology with respect to nucleic acids ranging from 70% to 88% with the HSERV9 sequence referenced X57147 and M37638 in the Genbank® data base. Four "consensus" nucleic acid sequences representative of different quasi-species of a possibly exogenous retrovirus MSRV-1B, or of different subfamilies of an endogenous retrovirus MSRV-1B, have been defined. These representative consensus sequences are presented in Figure 2, with the translation into amino acids. A functional reading frame exists for each subfamily of these MSRV-1B sequences, and it can be seen that the functional open reading frame corresponds in each instance to the amino acid sequence appearing on the second line under the nucleic acid sequence. The general

consensus of the MSRV-1B sequence, identified by SEQ ID NO:7 and obtained by this PCR technique in the "pol" region, is presented in Figure 1.

The second type of sequence representing the majority of the clones sequenced is represented by the sequence MSRV-2B presented in Figure 3 and identified by SEQ ID NO:11. The differences observed in the sequences corresponding to the PCR primers are explained by the use of degenerate primers in mixture form used under different technical conditions.

The MSRV-2B sequence (SEQ ID NO:11) is sufficiently divergent from the retroviral sequences already described in the data banks for it to be suggested that the sequence region in question belongs to a new infective agent, designated MSRV-2. This infective agent would, in principle, on the basis of the analysis of the first sequences obtained, be related to a retrovirus but, in view of the technique used for obtaining this sequence, it could also be a DNA virus whose genome codes for an enzyme which incidentally possesses reverse transcriptase activity, as is the case, for example, with the hepatitis B virus, HBV (12). Furthermore, the random nature of the degenerate primers used for this PCR amplification technique may very well have permitted, as a result of unforeseen sequence homologies or of conserved sites in the gene for a related enzyme, the amplification of a nucleic acid originating from a prokaryotic or eukaryotic pathogenic and/or coinfective agent (protist).

**EXAMPLE 2: OBTAINING CLONES DESIGNATED MSRV-1B AND MSRV-2B, DEFINING A FAMILY MSRV-1 and MSRV-2, BY "NESTED" PCR AMPLIFICATION OF THE CONSERVED POL REGIONS OF RETROVIRUSES ON PREPARATIONS OF B LYMPHOCYTES FROM A NEW CASE OF MS**

The same PCR technique, modified according to the technique of Shih (12), was used to amplify and

sequence the RNA nucleic acid material present in a purified fraction of virions at the peak of "LM7-like" reverse transcriptase activity on a sucrose gradient according to the technique described by H. Perron (13), and according to the protocols mentioned in Example 1, from a spontaneous lymphoblastoid line obtained by self-immortalization in culture of B lymphocytes from an MS patient who was seropositive for the Epstein-Barr virus (EBV), after setting up the blood lymphoid cells in culture in a suitable culture medium containing a suitable concentration of cyclosporin A. A representation of the reverse transcriptase activity in the sucrose fractions taken from a purification gradient of the virions produced by this line is presented in Figure 4. Similarly, the culture supernatants of a B line obtained under the same conditions from a control free from MS were treated under the same conditions, and the assay of reverse transcriptase activity in the sucrose gradient fractions proved negative throughout (background), and is presented in Figure 5. Fraction 3 of the gradient corresponding to the MS B line and the same fraction without reverse transcriptase activity of the non-MS control gradient were analysed by the same RT-PCR technique as before, derived from Shih (12), followed by the same steps of cloning and sequencing as described in Example 1.

It is particularly noteworthy that the MSRV-1 and MSRV-2 type sequences are to be found only in the material associated with a peak of "LM7-like" reverse transcriptase activity originating from the MS B lymphoblastoid line. These sequences were not to be found with the material from the control (non-MS) B lymphoblastoid line in 26 recombinant clones taken at random. Only Mo-MuLV type contaminant sequences, originating from the commercial reverse transcriptase used for the cDNA synthesis step, and sequences without any particular retroviral analogy were to be found in this control, as a

result of the "consensus" amplification of homologous polymerase sequences which is produced by this PCR technique. Furthermore, the absence of a concentrated target which competes for the amplification reaction in the control sample permits the amplification of dilute contaminants. The difference in results is manifestly highly significant (chi-squared,  $p < 0.001$ ).

EXAMPLE 3: OBTAINING A CLONE PSJ17, DEFINING A RETROVIRUS MSRV-1, BY REACTION OF ENDOGENOUS REVERSE TRANSCRIPTASE WITH A VIRION PREPARATION ORIGINATING FROM THE PLI-2 LINE

This approach is directed towards obtaining reverse-transcribed DNA sequences from the supposedly retroviral RNA in the isolate using the reverse transcriptase activity present in this same isolate. This reverse transcriptase activity can theoretically function only in the presence of a retroviral RNA linked to a primer tRNA or hybridized with short strands of DNA already reverse-transcribed in the retroviral particles (16). Thus, the obtaining of specific retroviral sequences in a material contaminated with cellular nucleic acids was optimized according to these authors by means of the specific enzymatic amplification of the portions of viral RNAs with a viral reverse transcriptase activity. To this end, the authors determined the particular physicochemical conditions under which this enzymatic activity of reverse transcription on RNAs contained in virions could be effective in vitro. These conditions correspond to the technical description of the protocols presented below (endogenous RT reaction, purification, cloning and sequencing).

The molecular approach consisted in using a preparation of concentrated but unpurified virion obtained from the culture supernatants of the PLI-2 line, prepared according to the following method: the culture

supernatants are collected twice weekly, precentrifuged at 10,000 rpm for 30 minutes to remove cell debris and then frozen at -80°C or used as they are for the following steps. The fresh or thawed supernatants are centrifuged on  
5 a cushion of 30% glycerol-PBS at 100,000 g (or 30,000 rpm in a type 45 T LKB-HITACHI rotor) for 2 h at 4°C. After removal of the supernatant, the sedimented pellet is taken up in a small volume of PBS and constitutes the fraction of concentrated but unpurified virion. This concentrated  
10 but unpurified viral sample was used to perform a so-called endogenous reverse transcription reaction, as described below.

A volume of 200 ml of virion purified according to the protocol described above, and containing a reverse  
15 transcriptase activity of approximately 1-5 million dpm, is thawed at 37°C until a liquid phase appears, and then placed on ice. A 5-fold concentrated buffer was prepared with the following components: 500 mM Tris-HCl pH 8.2; 75 mM NaCl; 25 mM MgCl<sub>2</sub>; 75 mM DTT and 0.10% NP 40; 100 ml  
20 of 5X buffer + 25 ml of a 100 mM solution of dATP + 25 ml of a 100 mM solution of dTTP + 25 ml of a 100 mM solution of dGTP + 25 ml of a 100 mM solution of dCTP + 100 ml of sterile distilled water + 200 ml of the virion suspension (RT activity of 5 million DPM) in PBS were mixed and  
25 incubated at 42°C for 3 hours. After this incubation, the reaction mixture is added directly to a buffered phenol/chloroform/isoamyl alcohol mixture (Sigma ref. P 3803); the aqueous phase is collected and one volume of sterile distilled water is added to the organic phase to  
30 re-extract the residual nucleic acid material. The collected aqueous phases are combined, and the nucleic acids contained are precipitated by adding 3M sodium acetate pH 5.2 to 1/10 volume + 2 volumes of ethanol + 1 ml of glycogen (Boehringer-Mannheim ref. 901 393) and  
35 placing the sample at -20°C for 4 h or overnight at +4°C. The precipitate obtained after centrifugation is then

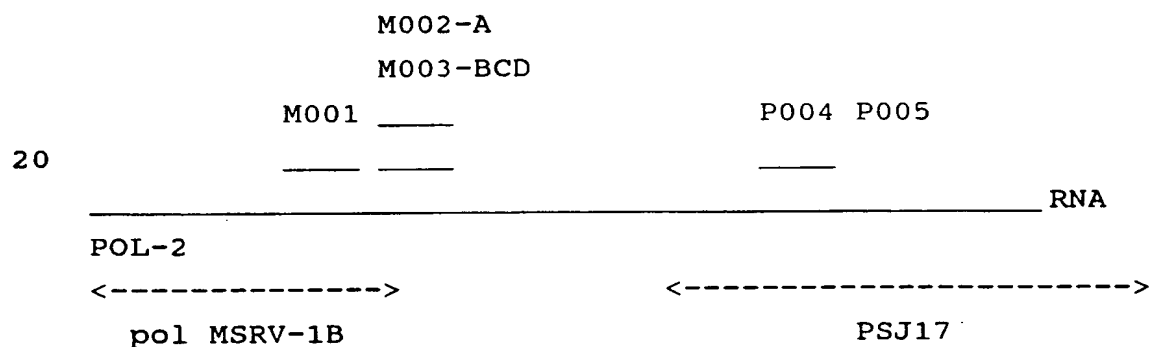
washed with 70% ethanol and resuspended in 60 ml of distilled water. The products of this reaction were then purified, cloned and sequenced according to the protocol which will now be described: blunt-ended DNAs with 5 unpaired adenines at the ends were generated: a "filling-in" reaction was first performed: 25 ml of the previously purified DNA solution were mixed with 2 ml of a 2.5 mM solution containing, in equimolar amounts, dATP + dGTP + dTTP + dCTP/1 ml of T4 DNA polymerase (Boehringer-Mannheim 10 ref. 1004 786) / 5 ml of 10X "incubation buffer for restriction enzyme" (Boehringer-Mannheim ref. 1417 975) / 1 ml of a 1% bovine serum albumin solution / 16 ml of sterile distilled water. This mixture was incubated for 20 minutes at 11°C. 50 ml of TE buffer and 1 ml of 15 glycogen (Boehringer-Mannheim ref. 901 393) were added thereto before extraction of the nucleic acids with phenol/chloroform/isoamyl alcohol (Sigma ref. P 3803) and precipitation with sodium acetate as described above. The DNA precipitated after centrifugation is resuspended in 20 10 ml of 10 mM Tris buffer pH 7.5. 5 ml of this suspension were then mixed with 20 ml of 5X Taq buffer, 20 ml of 5 mM dATP, 1 ml (5U) of Taq DNA polymerase (Amplitaq™) and 54 ml of sterile distilled water. This mixture is incubated for 2 h at 75°C with a film of oil on the 25 surface of the solution. The DNA suspended in the aqueous solution drawn off under the film of oil after incubation is precipitated as described above and resuspended in 2 ml of sterile distilled water. The DNA obtained was inserted into a plasmid using the TA Cloning™ kit. The 2 ml of DNA 30 solution were mixed with 5 ml of sterile distilled water, 1 ml of a 10-fold concentrated ligation buffer "10X LIGATION BUFFER", 2 ml of "pCR™ VECTOR" (25 ng/ml) and 1 ml of "TA DNA LIGASE". This mixture was incubated overnight at 12°C. The following steps were carried out 35 according to the instructions of the TA Cloning™ kit (British Biotechnology). At the end of the procedure, the

white colonies of recombinant bacteria (white) were picked out in order to be cultured and to permit extraction of the plasmids incorporated according to the so-called "miniprep" procedure (17). The plasmid preparation from each recombinant colony was cut with a suitable restriction enzyme and analysed on agarose gel. Plasmids possessing an insert detected under UV light after staining the gel with ethidium bromide were selected for sequencing of the insert, after hybridization with a primer complementary to the Sp6 promoter present on the cloning plasmid of the TA cloning™ kit. The reaction prior to sequencing was then performed according to the method recommended for the use of the sequencing kit "Prism ready reaction kit dye deoxyterminator cycle sequencing kit" (Applied Biosystems, ref. 401384), and automatic sequencing was carried out with an Applied Biosystems "Automatic Sequencer, model 373 A" apparatus according to the manufacturer's instructions.

Discriminating analysis on the computerized data banks of the sequences cloned from the DNA fragments present in the reaction mixture enabled a retroviral type sequence to be revealed. The corresponding clone PSJ17 was completely sequenced, and the sequence obtained, presented in Figure 6 and identified by SEQ ID NO:9, was analysed using the "Geneworks®" software on the updated "Genebank™" data banks. An identical sequence already described could not be found by analysis of the data banks. Only a partial homology with some known retroviral elements was to be found. The most useful relative homology relates to an endogenous retrovirus designated ERV-9, or HSERV-9, according to the references (18).

**EXAMPLE 4: PCR AMPLIFICATION OF THE NUCLEIC ACID SEQUENCE CONTAINED BETWEEN THE 5' REGION DEFINED BY THE CLONE "POL MSRV-1B" AND THE 3' REGION DEFINED BY THE CLONE PSJ17**

5           Five oligonucleotides, M001, M002-A, M003-BCD, P004 and P005, were defined in order to amplify the RNA originating from purified POL-2 virions. Control reactions were performed so as to check for the presence of contaminants (reaction with water). The amplification  
10 consists of an RT-PCR step according to the protocol described in Example 2, followed by a "nested" PCR according to the PCR protocol described in the document EP-A-0,569,272. In the first RT-PCR cycle, the primers M001 and P004 or P005 are used. In the second PCR cycle,  
15 the primers M002-A or M003-BCD and the primer P004 are used. The primers are positioned as follows:



25

Their composition is:

primer M001: GGTCITICCAIGG (SEQ ID NO:20)  
 primer M002-A: TTAGGGATAGCCCTCATCTCT (SEQ ID NO:21)  
 primer M003-BCD: TCAGGGATAGCCCCCATCTAT (SEQ ID NO:22)  
 30 primer P004: AACCCCTTGCCACTACATCAATTT (SEQ ID NO:23)  
 primer P005: GCGTAAGGACTCCTAGAGCTATT (SEQ ID NO:24)

The "nested" amplification product obtained, and designated M003-P004, is presented in Figure 7, and corresponds to the sequence SEQ ID NO:8.

35



**EXAMPLE 5: AMPLIFICATION AND CLONING OF A PORTION OF THE MSRV-1 RETROVIRAL GENOME USING A SEQUENCE ALREADY IDENTIFIED, IN A SAMPLE OF VIRUS PURIFIED AT THE PEAK OF REVERSE TRANSCRIPTASE ACTIVITY**

5 A PCR technique derived from the technique published by Frohman (19) was used. The technique derived makes it possible, using a specific primer at the 3' end of the genome to be amplified, to elongate the sequence towards the 5' region of the genome to be analysed. This  
10 technical variant is described in the documentation of the firm "Clontech Laboratories Inc.", (Palo-Alto California, USA) supplied with its product "5'-AmpliFINDER™ RACE Kit", which was used on a fraction of virion purified as described above.

15 The specific 3' primers used in the kit protocol for the synthesis of the cDNA and the PCR amplification are, respectively, complementary to the following MSRV-1 sequences:

20 cDNA:TCATCCATGTACCGAAGG (SEQ ID NO:25)  
amplification :ATGGGGTTCCCAAGTTCCT (SEQ ID NO:26)

The products originating from the PCR were obtained after purification on agarose gel according to conventional methods (17), and then resuspended in 10 ml  
25 of distilled water. Since one of the properties of Taq polymerase consists in adding an adenine at the 3' end of each of the two DNA strands, the DNA obtained was inserted directly into a plasmid using the TA Cloning™ kit (British Biotechnology). The 2 ml of DNA solution were  
30 mixed with 5 ml of sterile distilled water, 1 ml of a 10-fold concentrated ligation buffer "10X LIGATION BUFFER", 2 ml of "pCR™ VECTOR" (25 ng/ml) and 1 ml of "TA DNA LIGASE". This mixture was incubated overnight at 12°C. The  
35 following steps were carried out according to the instructions of the TA Cloning™ kit (British Biotechnology). At the end of the procedure, the white

colonies of recombinant bacteria (white) were picked out in order to be cultured and to permit extraction of the plasmids incorporated according to the so-called "mini-prep" procedure (17). The plasmid preparation from each recombinant colony was cut with a suitable restriction enzyme and analysed on agarose gel. Plasmids possessing an insert detected under UV light after staining the gel with ethidium bromide were selected for sequencing of the insert, after hybridization with a primer complementary to the Sp6 promoter present on the cloning plasmid of the TA Cloning<sup>TM</sup> Kit. The reaction prior to sequencing was then performed according to the method recommended for the use of the sequencing kit "Prism ready reaction kit dye deoxyterminator cycle sequencing kit" (Applied Biosystems, ref. 401384), and automatic sequencing was carried out with an Applied Biosystems "Automatic Sequencer model 373 A" apparatus according to the manufacturer's instructions.

This technique was applied first to two fractions of virion purified as described below on sucrose from the "POL-2" isolate produced by the PLI-2 line on the one hand, and from the MS7PG isolate produced by the LM7PC line on the other hand. The culture supernatants are collected twice weekly, precentrifuged at 10,000 rpm for 30 minutes to remove cell debris and then frozen at -80°C or used as they are for the following steps. The fresh or thawed supernatants are centrifuged on a cushion of 30% glycerol-PBS at 100,000 g (or 30,000 rpm in a type 45 T LKB-HITACHI rotor) for 2 h at 4°C. After removal of the supernatant, the sedimented pellet is taken up in a small volume of PBS and constitutes the fraction of concentrated but unpurified virions. The concentrated virus is then applied to a sucrose gradient in sterile PBS buffer (15 to 50% weight/weight) and ultracentrifuged at 35,000 rpm (100,000 g) for 12 h at +4°C in a swing-out rotor. 10 fractions are collected, and 20 ml are withdrawn from

each fraction after homogenization to assay the reverse transcriptase activity therein according to the technique described by H. Perron (3). The fractions containing the peak of "LM7-like" RT activity are then diluted in sterile  
5 PBS buffer and ultracentrifuged for one hour at 35,000 rpm (100,000 g) to sediment the viral particles. The pellet of purified virion thereby obtained is then taken up in a small volume of a buffer which is appropriate for the extraction of RNA. The cDNA synthesis reaction mentioned  
10 above is carried out on this RNA extracted from purified extracellular virion. PCR amplification according to the technique mentioned above enabled the clone F1-11 to be obtained, whose sequence, identified by SEQ ID NO:2, is presented in Figure 8.

15 This clone makes it possible to define, with the different clones previously sequenced, a region of considerable length (1.2 kb) representative of the "pol" gene of the MSRV-1 retrovirus, as presented in Figure 9. This sequence, designated SEQ ID NO:1, is reconstituted  
20 from different clones overlapping one another at their ends, correcting the artefacts associated with the primers and with the amplification or cloning techniques which would artificially interrupt the reading frame of the whole. This sequence will be identified below under the  
25 designation "MSRV-1 pol\* region". Its degree of homology with the HSERV-9 sequence is shown in Figure 12.

In Figure 9, the potential reading frame with its translation into amino acids is presented below the nucleic acid sequence.

30

**EXAMPLE 6: DETECTION OF SPECIFIC MSRV-1 and MSRV-2 SEQUENCES IN DIFFERENT SAMPLES OF PLASMA ORIGINATING FROM PATIENTS SUFFERING FROM MS OR FROM CONTROLS**

35 A PCR technique was used to detect the MSRV-1 and MSRV-2 genomes in plasmas obtained after taking blood

samples from patients suffering from MS and from non-MS controls onto EDTA.

Extraction of the RNAs from plasma was performed according to the technique described by P. Chomzynski (20), after adding one volume of buffer containing guanidinium thiocyanate to 1 ml of plasma stored frozen at -80°C after collection.

For MSRV-2, the PCR was performed under the same conditions and with the following primers:

- 10       - 5' primer, identified by SEQ ID NO:14  
          5' GTAGTTCGATGTAGAAAGCG 3';
- 3' primer, identified by SEQ ID NO:15  
          5' GCATCCGGCAACTGCACG 3'.

However, similar results were also obtained with the following PCR primers in two successive amplifications by "nested" PCR on samples of nucleic acids not treated with DNase.

The primers used for this first step of 40 cycles with a hybridization temperature of 48°C are the following:

- 20       - 5' primer, identified by SEQ ID NO:27  
          5' GCCGATATCACCCGCCATGG 3', corresponding to a  
          5' MSRV-2 PCR primer, for a first PCR on samples from  
          patients,
- 25       - 3' primer, identified by SEQ ID NO:28  
          5' GCATCCGGCAACTGCACG 3', corresponding to a 3'  
          MSRV-2 PCR primer, for a first PCR on samples from  
          patients.

After this step, 10 ml of the amplification product are taken and used to carry out a second, so-called "nested" PCR amplification with primers located within the region already amplified. This second step takes place over 35 cycles, with a primer hybridization ("annealing") temperature of 50°C. The reaction volume is 100 ml.

The primers used for this second step are the following:

- 5' primer, identified by SEQ ID NO:29

5' CGCGATGCTGGTTGGAGAGC 3', corresponding to a

5' MSR-2 PCR primer, for a nested PCR on samples from patients,

- 3' primer, identified by SEQ ID NO:30

5' TCTCCACTCCGAATATTCCG 3', corresponding to a

3' MSR-2 PCR primer, for a nested PCR on samples from patients.

For MSR-1, the amplification was performed in two steps. Furthermore, the nucleic acid sample is treated beforehand with DNase, and a control PCR without RT (AMV reverse transcriptase) is performed on the two amplification steps so as to verify that the RT-PCR amplification comes exclusively from the MSR-1 RNA. In the event of a positive control without RT, the initial aliquot sample of RNA is again treated with DNase and amplified again.

The protocol for treatment with DNase lacking RNase activity is as follows: the extracted RNA is aliquoted in the presence of "RNase inhibitor" (Boehringer-Mannheim) in water treated with DEPC at a final concentration of 1 mg in 10 ml; to these 10 ml, 1 ml of "RNase-free DNase" (Boehringer-Mannheim) and 1.2 ml of pH 5 buffer containing 0.1 M/l sodium acetate and 5 mM/l  $\text{MgSO}_4$  is added; the mixture is incubated for 15 min at 20°C and brought to 95°C for 1.5 min in a "thermocycler".

The first MSR-1 RT-PCR step is performed according to a variant of the RNA amplification method as described in Patent Application No. EP-A-0,569,272. In particular, the cDNA synthesis step is performed at 42°C for one hour; the PCR amplification takes place over 40 cycles, with a primer hybridization ("annealing") temperature of 53°C. The reaction volume is 100 µl.

The primers used for this first step are the following:

- 5' primer, identified by SEQ ID NO:16  
5' AGGAGTAAGGAAACCCAACGGAC 3';
- 5     - 3' primer, identified by SEQ ID NO:17  
5' TAAGAGTTGCACAAGTGCG 3'.

After this step, 10 ml of the amplification product are taken and used to carry out a second, so-called "nested" PCR amplification with primers located within the region already amplified. This second step takes place over 35 cycles, with a primer hybridization ("annealing") temperature of 53°C. The reaction volume is 100 ml.

The primers used for this second step are the following:

- 5' primer, identified by SEQ ID NO:18  
5' TCAGGGATAGCCCCATCTAT 3';
- 3' primer, identified by SEQ ID NO:19  
5' AACCCCTTTGCCACTACATCAATTT 3'.

Figures 10 and 11 present the results of PCR in the form of photographs under ultraviolet light of ethidium bromide-impregnated agarose gels, in which an electrophoresis of the PCR amplification products applied separately to the different wells was performed.

The top photograph (Figure 10) shows the result of specific MSRV-2 amplification.

Well number 8 contains a mixture of DNA molecular weight markers, and wells 1 to 7 represent, in order, the products amplified from the total RNAs of plasmas originating from 4 healthy controls free from MS (wells 1 to 4) and from 3 patients suffering from MS at different stages of the disease (wells 5 to 7).

In this series, MSRV-2 nucleic acid material is detected in the plasma of one case of MS out of the 3 tested, and in none of the 4 control plasmas. Other

results obtained on more extensive series confirm these results.

The bottom photograph (Figure 11) shows the result of specific amplification by MSRV-1 "nested"

5 RT-PCR:

well No. 1 contains the PCR product produced with water alone, without the addition of AMV reverse transcriptase; well No. 2 contains the PCR product produced with water alone, with the addition of AMV  
10 reverse transcriptase; well number 3 contains a mixture of DNA molecular weight markers; wells 4 to 13 contain, in order, the products amplified from the total RNAs extracted from sucrose gradient fractions (collected in a downward direction), on which gradient a pellet of virion  
15 originating from a supernatant of a culture infected with MSRV-1 and MSRV-2 was centrifuged to equilibrium according to the protocol described by H. Perron (13); to well 14 nothing was applied; to wells 15 to 17, the amplified products of RNA extracted from plasmas originating from 3  
20 different patients suffering from MS at different stages of the disease were applied.

The MSRV-1 retroviral genome is indeed to be found in the sucrose gradient fraction containing the peak of reverse transcriptase activity measured according to  
25 the technique described by H. Perron (3), with a very strong intensity (fraction 5 of the gradient, placed in well No. 8). A slight amplification has taken place in the first fraction (well No. 4), probably corresponding to RNA released by lysed particles which floated at the surface  
30 of the gradient; similarly, aggregated debris has sedimented in the last fraction (tube bottom), carrying with it a few copies of the MSRV-1 genome which have given rise to an amplification of low intensity.

Of the 3 MS plasmas tested in this series, MSRV-  
35 1 RNA turned up in one case, producing a very intense amplification (well No. 17).

In this series, the MSRV-1 retroviral RNA genome, probably corresponding to particles of extracellular virus present in the plasma in extremely small numbers, was detected by "nested" RT-PCR in one case  
5 of MS out of the 3 tested. Other results obtained on more extensive series confirm these results.

Furthermore, the specificity of the sequences amplified by these PCR techniques may be verified and evaluated by the "ELOSAs" technique as described by  
10 F. Mallet (21) and in the document FR-A-2,663,040.

For MSRV-1, the products of the nested PCR described above may be tested in two ELOSA systems enabling a consensus A and a consensus B+C+D of MSRV-1 to be detected separately, corresponding to the subfamilies  
15 described in Example 1 and Figures 1 and 2. In effect, the sequences closely resembling the consensus B+C+D are to be found essentially in the RNA samples originating from MSRV-1 virions purified from cultures or amplified in extracellular biological fluids of MS patients, whereas  
20 the sequences closely resembling the consensus A are essentially to be found in normal human cellular DNA.

The ELOSA/MSRV-1 system for the capture and specific hybridization of the PCR products of the subfamily A uses a capture oligonucleotide cpV1A with an  
25 amine bond at the 5' end and a biotinylated detection oligonucleotide dpV1A having as their sequence, respectively:

- cpV1A identified by SEQ ID NO:31

5' GATCTAGGCCACTTCTCAGGTCCAGS 3', corresponding  
30 to the ELOSA capture oligonucleotide for the products of MSRV-1 nested PCR performed with the primers identified by SEQ ID NO:16 and SEQ ID NO:17, optionally followed by amplification with the primers identified by SEQ ID NO:18 and SEQ ID NO:19 on samples from patients;

35 - dpV1A identified by SEQ ID NO:32;



5' CATCTITTTGGICAGGCAITAGC 3', corresponding to the ELOSA capture oligonucleotide for the subfamily A of the products of MSRV-1 "nested" PCR performed with the primers identified by SEQ ID NO:16 and SEQ ID NO:17, optionally followed by amplification with the primers identified by SEQ ID NO:18 and SEQ ID NO:19 on samples from patients.

The ELOSA/MSRV-1 system for the capture and specific hybridization of the PCR products of the subfamily B+C+D uses the same biotinylated detection oligonucleotide dpV1A and a capture oligonucleotide cpV1B with an amine bond at the 5' end having as its sequence:

- dpV1B identified by SEQ ID NO:33

5' CTTGAGCCAGTTCTCATACCTGGA 3', corresponding to the ELOSA capture oligonucleotide for the subfamily B + C + D of the products of MSRV-1 "nested" PCR performed with the primers identified by SEQ ID NO:16 and SEQ ID NO:17, optionally followed by amplification with the primers identified by SEQ ID NO:18 and SEQ ID NO:19 on samples from patients.

This ELOSA detection system enabled it to be verified that none of the PCR products thus amplified from DNase-treated plasmas of MS patients contained a sequence of the subfamily A, and that all were positive with the consensus of the subfamilies B, C and D.

For MSRV-2, a similar ELOSA technique was evaluated on isolates originating from infected cell cultures, using the following PCR amplification primers,

- 5' primer, identified by SEQ ID NO:34

5' AGTGYTRCCMCARGGCGCTGAA 3', corresponding to a 5' MSRV-2 PCR primer, for PCR on samples from cultures,

- 3' primer, identified by SEQ ID NO:35

5' GMGGCCAGCAGSAKGTTCATCCA 3', corresponding to a 3' MSRV-2 PCR primer, for PCR on samples from cultures,

and the capture oligonucleotides with an amine bond at the 5' end cpV2 and the biotinylated detection oligonucleotide dpV2 having as their respective sequences:

- cpV2 identified by SEQ ID NO:36

5 5 GGATGCCGCCTATAGCCTCTAC 3', corresponding to an ELOSA capture oligonucleotide for the products of MSRV-2 PCR performed with the primers SEQ ID NO:34 and SEQ ID NO:35, or optionally with the degenerate primers defined by Shih (12).

10 - dpV2 identified by SEQ ID NO:37

5' AAGCCTATCGCGTGAGTTGCC 3', corresponding to an ELOSA detection oligonucleotide for the products of MSRV-2 PCR performed with the primers SEQ ID NO:34 and SEQ ID NO:35, or optionally with the degenerate primers  
15 defined by Shih (12)

This PCR amplification system with a pair of primers different from those which were described previously for amplification on the samples from patients made it possible to confirm the infection with MSRV-2 of in  
20 vitro cultures and of samples of nucleic acids used for the molecular biology studies.

All things considered, the first results of PCR detection of the genome of pathogenic and/or infective agents show that it is possible that free "virus" may  
25 circulate in the blood stream of patients in an acute, virulent phase, outside the nervous system. This is compatible with the almost invariable presence of "gaps" in the blood-brain barrier of patients in an active phase of MS.

30

#### EXAMPLE 7: OBTAINING SEQUENCES OF THE "env" GENE OF THE MSRV-1 RETROVIRAL GENOME

As has already been described in Example 5, a PCR technique derived from the technique published by  
35 Frohman (19) was used. The technique derived makes it possible, using a specific primer at the 3' end of the

genome to be amplified, to elongate the sequence towards the 5' region of the genome to be analysed. This technical variant is described in the documentation of "Clontech Laboratories Inc., (Palo-Alto California, USA) supplied  
5 with its product "5'-AmpliFINDER™ RACE Kit", which was used on a fraction of virion purified as described above.

In order to carry out an amplification of the 3' region of the MSRV-1 retroviral genome encompassing the region of the "env" gene, a study was carried out to  
10 determine a consensus sequence in the LTR regions of the same type as those of the defective endogenous retrovirus HSERV-9 (18, 24), with which the MSRV-1 retrovirus displays partial homologies.

The same specific 3' primer was used in the kit  
15 protocol for the synthesis of the cDNA and the PCR amplification; its sequence is as follows:

GTGCTGATTGGTGTATTTACAATCC (SEQ ID NO 45)

Synthesis of the complementary DNA (cDNA) and unidirectional PCR amplification with the above primer  
20 were carried out in one step according to the method described in Patent EP-A-0,569,272.

The products originating from the PCR were extracted after purification of agarose gel according to conventional methods (17), and then resuspended in 10 ml  
25 of distilled water. Since one of the properties of Taq polymerase consists in adding an adenine at the 3' end of each of the two DNA strands, the DNA obtained was inserted directly into a plasmid using the TA Cloning™ kit (British Biotechnology). The 2 ml of DNA solution were mixed with 5  
30 ml of sterile distilled water, 1 ml of a 10-fold concentrated ligation buffer "10X LIGATION BUFFER", 2 ml of "pCR™ VECTOR" (25 ng/ml) and 1 ml of "TA DNA LIGASE". This mixture was incubated overnight at 12°C. The following steps were carried out according to the  
35 instructions of the TA Cloning® kit (British Biotechnology). At the end of the procedure, the white colonies of

recombinant bacteria (white) were picked out in order to be cultured and to permit extraction of the plasmids incorporated according to the so-called "miniprep" procedure (17). The plasmid preparation from each recombinant colony was cut with a suitable restriction enzyme and analysed on agarose gel. Plasmids possessing an insert detected under UV light after staining the gel with ethidium bromide were selected for sequencing of the insert, after hybridization with a primer complementary to the Sp6 promoter present on the cloning plasmid of the TA Cloning™ Kit. The reaction prior to sequencing was then performed according to the method recommended for the use of the sequencing kit "Prism ready reaction kit dye deoxyterminator cycle sequencing kit" (Applied Biosystems, ref. 401384), and automatic sequencing was carried out with an Applied Biosystems "automatic sequencer, model 373 A" apparatus according to the manufacturer's instructions.

This technical approach was applied to a sample of virion concentrated as described below from a mixture of culture supernatants produced by B lymphoblastoid lines such as are described in Example 2, established from lymphocytes of patients suffering from MS and possessing reverse transcriptase activity which is detectable according to the technique described by Perron et al. (3): the culture supernatants are collected twice weekly, precentrifuged at 10,000 rpm for 30 minutes to remove cell debris and then frozen at -80°C or used as they are for the following steps. The fresh or thawed supernatants are centrifuged on a cushion of 30% glycerol-PBS at 100,000 g for 2 h at 4°C. After removal of the supernatant, the sedimented pellet constitutes the sample of concentrated but unpurified virions. The pellet thereby obtained is then taken up in a small volume of an appropriate buffer for the extraction of RNA. The cDNA synthesis reaction

mentioned above is carried out on this RNA extracted from concentrated extracellular virion.

RT-PCR amplification according to the technique mentioned above enabled the clone FBd3 to be obtained, whose sequence, identified by SEQ ID NO:46, is presented in Figure 13.

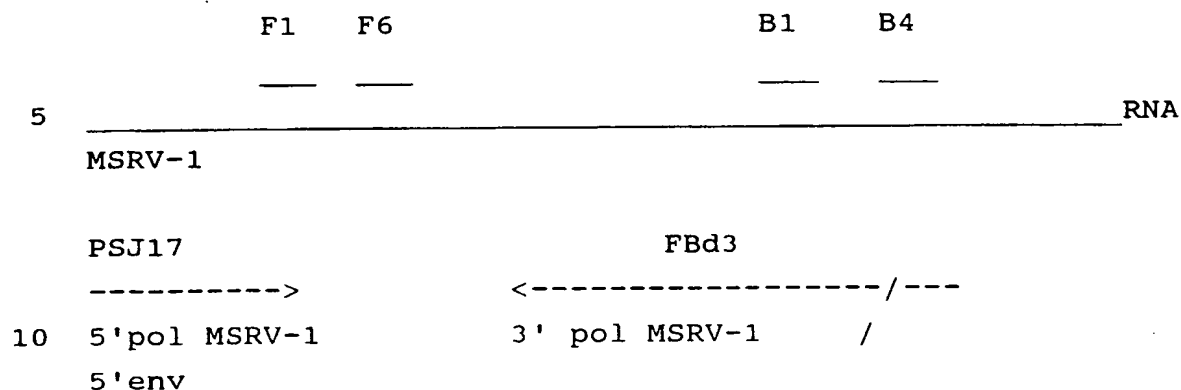
In Figure 14, the sequence homology between the clone FBd3 and the HSERV-9 retrovirus is shown on the matrix chart by a continuous line for any partial homology greater than or equal to 65%. It can be seen that there are homologies in the flanking regions of the clone (with the pol gene at the 5' end and with the env gene and then the LTR at the 3' end), but that the internal region is totally divergent and does not display any homology, even weak, with the "env" gene of HSERV9. Furthermore, it is apparent that the clone FBd3 contains a longer "env" region than the one which is described for the defective endogenous HSERV-9; it may thus be seen that the internal divergent region constitutes an "insert" between the regions of partial homology with the HSERV-9 defective genes.

#### EXAMPLE 8: AMPLIFICATION, CLONING AND SEQUENCING OF THE REGION OF THE MSRV-1 RETROVIRAL GENOME LOCATED BETWEEN THE CLONES PSJ17 AND FBd3

Four oligonucleotides, F1, B4, F6 and B1, were defined for amplifying RNA originating from concentrated virions of the strains POL2 and MS7PG. Control reactions were performed so as to check for the presence of contaminants (reaction with water). The amplification consists of a first step of RT-PCR according to the protocol described in Patent Application EP-A-0,569,272, followed by a second step of PCR performed on 10 ml of product of the first step with primers internal to the amplified first region ("nested" PCR). In the first RT-PCR cycle, the primers F1 and B4 are used. In the second PCR

56

cycle, the primers F6 and the primer B1 are used. The primers are positioned as follows:



Their composition is:

- primer F1: TGATGTGAACGGCATACTCACTG (SEQ ID NO:47)  
 15 primer B4: CCCAGAGGTTAGGAAGTCCCTTTC (SEQ ID NO 48)  
 primer F6: GCTAAAGGAGACTTGTGGTTGTCAG (SEQ ID NO 49)  
 primer B1: CAACATGGGCATTTTCGGATTAG (SEQ ID NO 50)

The product of "nested" amplification obtained and designated "t pol" is presented in Figure 15, and  
 20 corresponds to the sequence SEQ ID NO:51.

**EXAMPLE 9: OBTAINING NEW SEQUENCES, EXPRESSED AS RNA IN CELLS IN CULTURE PRODUCING MSRV-1, AND COMPRISING AN "env" REGION OF THE MSRV-1 RETROVIRAL GENOME**

25 A library of cDNA was produced according to the procedure described by the manufacturer of the "cDNA synthesis module, cDNA rapid adaptator ligation module, cDNA rapid cloning module and lambda gt10 in vitro packaging module" kits (Amersham, ref RPN1256Y/Z, RPN1712,  
 30 RPN1713, RPN1717, N334Z), from the messenger RNA extracted from cells of a B lymphoblastoid line such as is described in Example 2, established from the lymphocytes of a patient suffering from MS and possessing reverse transcriptase activity which is detectable according to  
 35 the technique described by Perron et al. (3).

Oligonucleotides were defined for amplifying the cDNA cloned into the nucleic acid library between the 3' region of the clone PSJ17 (pol) and the 5'(LTR) region of the clone FBd3. Control reactions were performed so as to check for the presence of contaminants (reaction with water). PCR reactions performed on the nucleic acids cloned into the library with different pairs of primers enabled a series of clones linking pol sequences to the MSRV-1 type env or LTR sequences to be amplified.

Two clones are representative of the sequences obtained in the cellular cDNA library:

- the clone JLBc1, whose sequence SEQ ID NO:52 is presented in Figure 16;
- the clone JLBc2, whose sequence SEQ ID NO:53 is presented in Figure 17.

The sequences of the clones JLBc1 and JLBc2 are homologous to that of the clone FBd3, as is apparent in Figures 18 and 19. The homology between the clone JLBc1 and the clone JLBc2 is shown in Figure 20.

The homologies between the clones JLBc1 and JLBc2 on the one hand and the HSERV9 sequence on the other hand are presented, respectively, in Figures 21 and 22.

It will be noted that the region of homology between JLB1, JLB2 and FBd3 comprises, with a few sequence and size variations of the "insert", the additional sequence absent ("inserted") in the HSERV-9 env sequence, as described in Example 8.

It will also be noted that the cloned "pol" region is very homologous to HSERV-9, does not possess a reading frame (bearing in mind the sequence errors induced by the techniques used, including even the automatic sequencer) and diverges from the MSRV-1 sequences obtained from virions. In view of the fact that these sequences were cloned from the RNA of cells expressing MSRV-1 particles, it is probable that they originate from endogenous retroviral elements related to the ERV9 family;

this is all the more likely for the fact that the pol and env genes are present on the same RNA which is clearly not the MSRV-1 genomic RNA. Some of these ERV9 elements possess functional LTRs which can be activated by replicative viruses coding for homologous or heterologous transactivators. Under these conditions, the relationship between MSRV-1 and HSERV-9 makes probable the transactivation of the defective (or otherwise) endogenous ERV9 elements by homologous, or even identical, MSRV-1 transactivating proteins.

Such a phenomenon may induce a viral interference between the expression of MSRV-1 and the related endogenous elements. Such an interference generally leads to a so-called "defective-interfering" expression, some features of which were to be found in the MSRV-1-infected cultures studied. Furthermore, such a phenomenon does not lack generation of the expression of polypeptides, or even of endogenous retroviral proteins which are not necessarily tolerated by the immune system. Such a scheme of aberrant expression of endogenous elements related to MSRV-1 and induced by the latter is liable to multiply the aberrant antigens, and hence to contribute to the induction of autoimmune processes such as are observed in MS.

It is, however, essential to note that the clones JLBc1 and JLBc2 differ from the ERV9 or HSERV9 sequence already described, in that they possess a longer env region comprising an additional region totally divergent from ERV9. Their kinship with the endogenous ERV9 family may hence be defined, but they clearly constitute novel elements never hitherto described. In effect, interrogation of the data banks of nucleic acid sequences available in version No. 15 (1995) of the "Entrez" software (NCBI, NIH, Bethesda, USA) did not enable a known homologous sequence in the env region of these clones to be identified.



**EXAMPLE 10: OBTAINING SEQUENCES LOCATED IN THE  
5' pol AND 3' gag REGION OF THE MSRV-1 RETROVIRAL GENOME**

As has already been described in Example 5, a  
5 PCR technique derived from the technique published by  
Frohman (19) was used. The technique derived makes it  
possible, using a specific primer at the 3' end of the  
genome to be amplified, to elongate the sequence towards  
the 5' region of the genome to be analysed. This technical  
10 variant is described in the documentation of the firm  
Clontech Laboratories Inc., (Palo-Alto California, USA)  
supplied with its product "5'-AmpliFINDER™ RACE Kit",  
which was used on a fraction of virion purified as  
described above.

15 In order to carry out an amplification of the 5'  
region of the MSRV-1 retroviral genome starting from the  
pol sequence already sequenced (clone F11-1) and extending  
towards the gag gene, MSRV-1 specific primers were  
defined.

20 The specific 3' primers used in the kit protocol  
for the synthesis of the cDNA and the PCR amplification  
are, respectively, complementary to the following MSRV-1  
sequences:

25 cDNA: (SEQ ID NO:54)  
CCTGAGTTCTTGCACTAACCC  
amplification: (SEQ ID NO:55)  
GTCCGTTGGGTTTCCTTACTCCT

The products originating from the PCR were  
extracted after purification on agarose gel according to  
30 conventional methods (17), and then resuspended in 10 ml  
of distilled water. Since one of the properties of Taq  
polymerase consists in adding an adenine at the 3' end of  
each of the two DNA strands, the DNA obtained was inserted  
directly into a plasmid using the TA Cloning™ kit (British  
35 Biotechnology). The 2 ml of DNA solution were mixed with 5  
ml of sterile distilled water, 1 ml of a 10-fold

concentrated ligation buffer "10X LIGATION BUFFER", 2 ml of "pCR™ VECTOR" (25 ng/ml) and 1 ml of "TA DNA LIGASE". This mixture was incubated overnight at 12°C. The following steps were carried out according to the instructions of the TA Cloning® kit (British Biotechnology). At the end of the procedure, the white colonies of recombinant bacteria (white) were picked out in order to be cultured and to permit extraction of the plasmids incorporated according to the so-called "miniprep" procedure (17). The plasmid preparation from each recombinant colony was cut with a suitable restriction enzyme and analysed on agarose gel. Plasmids possessing an insert detected under UV light after staining the gel with ethidium bromide were selected for sequencing of the insert, after hybridization with a primer complementary to the Sp6 promoter present on the cloning plasmid of the TA Cloning™ Kit. The reaction prior to sequencing was then performed according to the method recommended for the use of the sequencing kit "Prism ready reaction kit dye deoxyterminator cycle sequencing kit" (Applied Biosystems, ref. 401384), and automatic sequencing was carried out with an Applied Biosystems "automatic sequencer model 373 A" apparatus according to the manufacturer's instructions.

This technical approach was applied to a sample of virion concentrated as described below from a mixture of culture supernatants produced by B lymphoblastoid lines such as are described in Example 2, established from lymphocytes of patients suffering from MS and possessing reverse transcriptase activity which is detectable according to the technique described by Perron et al. (3): the culture supernatants are collected twice weekly, precentrifuged at 10,000 rpm for 30 minutes to remove cell debris and then frozen at -80°C or used as they are for the following steps. The fresh or thawed supernatants are centrifuged on a cushion of 30% glycerol-PBS at 100,000 g

for 2 h at 4°C. After removal of the supernatant, the sedimented pellet constitutes the sample of concentrated but unpurified virions. The pellet thereby obtained is then taken up in a small volume of an appropriate buffer  
5 for the extraction of RNA. The cDNA synthesis reaction mentioned above is carried out on this RNA extracted from concentrated extracellular virion.

RT-PCR amplification according to the technique mentioned above enabled the clone GM3 to be obtained,  
10 whose sequence, identified by SEQ ID NO 56, is presented in Figure 23.

In Figure 24, the sequence homology between the clone GMP3 and the HSERV-9 retrovirus is shown on the matrix chart by a continuous line, for any partial  
15 homology greater than or equal to 65%.

In summary, Figure 25 shows the localization of the different clones studied above, relative to the known ERV9 genome. In Figure 25, since the MSRV-1 env region is longer than the reference ERV9 env gene, the additional  
20 region is shown above the point of insertion according to a "V", on the understanding that the inserted material displays a sequence and size variability between the clones shown (JLBc1, JLBc2, FBd3). And Figure 26 shows the position of different clones studied in the MSRV-1 pol\*  
25 region.

By means of the clone GM3 described above, a possible reading frame could be defined, covering the whole of the pol gene, referenced according to SEQ ID NO:57, shown in the successive Figures 27a to 27c.

30

#### **EXAMPLE 11: DETECTION OF ANTI-MSRV-1 SPECIFIC ANTIBODIES IN HUMAN SERUM**

Identification of the sequence of the pol gene of the MSRV-1 retrovirus and of an open reading frame of  
35 this gene enabled the amino acid sequence SEQ ID NO:39 of

a region of the said gene, referenced SEQ ID NO:40, to be determined (see Figure 28).

Different synthetic peptides corresponding to fragments of the protein sequence of MSRV-1 reverse transcriptase encoded by the pol gene were tested for  
5 their antigenic specificity with respect to sera of patients suffering from MS and of healthy controls.

The peptides were synthesized chemically by solid-phase synthesis according to the Merrifield technique (Barany G, and Merrifield R.B, 1980, In the  
10 Peptides, 2, 1-284, Gross E and Meienhofer J, Eds., Academic Press, New York). The practical details are those described below.

a) Peptide synthesis:

15 The peptides were synthesized on a phenylacetamidomethyl (PAM)/polystyrene/divinylbenzene resin (Applied Biosystems, Inc. Foster City, CA), using an "Applied Biosystems 430A" automatic synthesizer. The amino acids are coupled in the form of hydroxybenzotriazole  
20 (HOBT) esters. The amino acids used are obtained from Novabiochem (Läufelfingen, Switzerland) or Bachem (Bubendorf, Switzerland).

The chemical synthesis was performed using a double coupling protocol with N-methylpyrrolidone (NMP) as  
25 solvent. The peptides were cut from the resin, as well as the side-chain protective groups, simultaneously, using hydrofluoric acid (HF) in a suitable apparatus (type I cleavage apparatus, Peptide Institute, Osaka, Japan).

For 1 g of peptidyl resin, 10 ml of HF, 1 ml of anisole and 1 ml of dimethyl sulphide 5DMS are used. The  
30 mixture is stirred for 45 minutes at -2°C. The HF is then evaporated off under vacuum. After intensive washes with ether, the peptide is eluted from the resin with 10% acetic acid and then lyophilized.

35 The peptides are purified by preparative high performance liquid chromatography on a VYDAC C18 type

column (250 x 21 mm) (The Separation Group, Hesperia, CA, USA). Elution is carried out with an acetonitrile gradient at a flow rate of 22 ml/min. The fractions collected are monitored by an elution under isocratic conditions on a  
5 VYDAC® C18 analytical column (250 x 4.6 mm) at a flow rate of 1 ml/min. Fractions having the same retention time are pooled and lyophilized. The preponderant fraction is then analysed by analytical high performance liquid chromatography with the system described above. The  
10 peptide which is considered to be of acceptable purity manifests itself in a single peak representing not less than 95% of the chromatogram.

The purified peptides are then analysed with the object of monitoring their amino acid composition, using  
15 an Applied Biosystems 420H automatic amino acid analyser. Measurement of the (average) chemical molecular mass of the peptides is obtained using LSIMS mass spectrometry in the positive ion mode on a VG. ZAB.ZSEQ double focusing instrument connected to a DEC-VAX 2000 acquisition system  
20 (VG analytical Ltd, Manchester, England).

The reactivity of the different peptides was tested against sera of patients suffering from MS and against sera of healthy controls. This enabled a peptide designated POL2B to be selected, whose sequence is shown  
25 in Figure 28 in the identifier SEQ ID NO:39, below, encoded by the pol gene of MSRV-1 (nucleotides 181 to 330).

b) Antigenic properties:

The antigenic properties of the POL2B peptide  
30 were demonstrated according to the ELISA protocol described below.

The lyophilized POL2B peptide was dissolved in sterile distilled water at a concentration of 1 mg/ml. This stock solution was aliquoted and kept at +4°C for use  
35 over a fortnight, or frozen at -20°C for use within 2 months. An aliquot is diluted in PBS (phosphate buffered

saline) solution so as to obtain a final peptide concentration of 1 microgram/ml. 100 microlitres of this dilution are placed in each well of microtitration plates ("high-binding" plastic, COSTAR ref: 3590). The plates are covered with a "plate-sealer" type adhesive and kept overnight at +4°C for the phase of adsorption of the peptide to the plastic. The adhesive is removed and the plates are washed three times with a volume of 300 microlitres of a solution A (1X PBS, 0.05% Tween 20®), then inverted over an absorbent tissue. The plates thus drained are filled with 200 microlitres per well of a solution B (solution A + 10% of goat serum), then covered with an adhesive and incubated for 45 minutes to 1 hour at 37°C. The plates are then washed three times with the solution A as described above.

The test serum samples are diluted beforehand to 1/50 in the solution B, and 100 microlitres of each dilute test serum are placed in the wells of each microtitration plate. A negative control is placed in one well of each plate, in the form of 100 microlitres of buffer B. The plates covered with an adhesive are then incubated for 1 to 3 hours at 37°C. The plates are then washed three times with the solution A as described above. In parallel, a peroxidase-labelled goat antibody directed against human IgG (Sigma Immunochemicals ref. A6029) or IgM (Cappel ref. 55228) is diluted in the solution B (dilution 1/5000 for the anti-IgG and 1/1000 for the anti-IgM). 100 microlitres of the appropriate dilution of the labelled antibody are then placed in each well of the microtitration plates, and the plates covered with an adhesive are incubated for 1 to 2 hours at 37°C. A further washing of the plates is then performed as described above. In parallel, the peroxidase substrate is prepared according to the directions of the "Sigma fast OPD kit" (Sigma Immunochemicals, ref. P9187). 100 microlitres of substrate solution are placed in each

well, and the plates are placed protected from light for 20 to 30 minutes at room temperature.

When the colour reaction has stabilized, the plates are placed immediately in an ELISA plate spectrophotometric reader, and the optical density (OD) of each well is read at a wavelength of 492 nm. Alternatively, 30 microlitres of 1N HCl are placed in each well to stop the reaction, and the plates are read in the spectrophotometer within 24 hours.

The serological samples are introduced in duplicate or in triplicate, and the optical density (OD) corresponding to the serum tested is calculated by taking the mean of the OD values obtained for the same sample at the same dilution.

The net OD of each serum corresponds to the mean OD of the serum minus the mean OD of the negative control (solution B: PBS, 0.05% Tween 20®, 10% goat serum).

c) Detection of anti-MSRV-1 IgG antibodies by ELISA:

The technique described above was used with the POLB2 peptide to test for the presence of anti-MSRV-1 specific IgG antibodies in the serum of 29 patients for whom a definite or probable diagnosis of MS was established according to the criteria of Poser (23), and of 32 healthy controls (blood donors).

Figure 29 shows the results for each serum tested with an anti-IgG antibody. Each vertical bar represents the net optical density (OD at 492 nm) of a serum tested. The ordinate axis gives the net OD at the top of the vertical bars. The first 29 vertical bars lying to the left of the vertical broken line represent the sera of 29 cases of MS tested, and the 32 vertical bars lying to the right of the vertical broken line represent the sera of 32 healthy controls (blood donors).

The mean of the net OD values for the MS sera tested is 0.62. The diagram enables 5 controls to be

revealed whose net OD rises above the grouped values of the control population. These values may represent the presence of specific IgGs in symptomless seropositive patients. Two methods were hence evaluated in order to  
5 determine the statistical threshold of positivity of the test.

The mean of the net OD values for the controls, including the controls with high net OD values, is 0.36. Without the 5 controls whose net OD values are greater  
10 than or equal to 0.5, the mean of the "negative" controls is 0.33. The standard deviation of the negative controls is 0.10. A theoretical threshold of positivity may be calculated according to the formula:  
threshold value (mean of the net OD values of the  
15 seronegative controls) + (2 or 3 x standard deviation of the net OD values of the seronegative controls).

In the first case, there are considered to be symptomless seropositives, and the threshold value is equal to  $0.33 + (2 \times 0.10) = 0.53$ . The negative results  
20 represent a non-specific "background" of the presence of antibodies directed specifically against an epitope of the peptide.

In the second case, if the set of controls consisting of blood donors in apparent good health is  
25 taken as a reference basis, without excluding the sera which are, on the face of it, seropositive, the standard deviation of the "non-MS controls" is 0.116. The threshold value then becomes  $0.36 + (2 \times 0.116) = 0.59$ .

According to this analysis, the test is specific  
30 for MS. In this respect, it is seen that the test is specific for MS, since, as shown in Table 1, no control has a net OD above this threshold. In fact, this result reflects the fact that the antibody titres in patients suffering from MS are, for the most part, higher than in  
35 healthy controls who have been in contact with MSRV-1.



TABLE No. 1

	MS	CONTROLS
	0.681	0.3515
	1.0425	0.56
5	0.5675	0.3565
	0.63	0.449
	0.588	0.2825
	0.645	0.55
	0.6635	0.52
10	0.576	0.2535
	0.7765	0.55
	0.5745	0.51
	0.513	0.426
	0.4325	0.451
15	0.7255	0.227
	0.859	0.3905
	0.6435	0.265
	0.5795	0.4295
	0.8655	0.291
20	0.671	0.347
	0.596	0.4495
	0.662	0.3725
	0.602	0.181
	0.525	0.2725
25	0.53	0.426
	0.565	0.1915
	0.517	0.222
	0.607	0.395
	0.3705	0.34
30	0.397	0.307
	0.4395	0.219
		0.491
		0.2265
		0.2605
35	MEAN 0.62	0.33
	STD DEV 0.14	0.10
	THRESHOLD VALUE	0.53

In accordance with the first method of calculation, and as shown in Figure 29 and in the corresponding Table 1, 26 of the 29 MS sera give a positive result (net OD greater than or equal to 0.50), indicating the presence of IgGs specifically directed against the POL2B peptide, hence against a portion of the reverse transcriptase enzyme of the MSRV-1 retrovirus encoded by its pol gene, and consequently against the MSRV-1 retrovirus. Thus, approximately 90% of the MS patients tested have reacted against an epitope carried by the POL2B peptide and possess circulating IgGs directed against the latter.

Five out of 32 blood donors in apparent good health show a positive result. Thus, it is apparent that approximately 15% of the symptomless population may have been in contact with an epitope carried by the POL2B peptide under conditions which have led to an active immunization which manifests itself in the persistence of specific serum IgGs. These conditions are compatible with an immunization against the MSRV-1 retrovirus reverse transcriptase during an infection with (and/or reactivation of) the MSRV-1 retrovirus. The absence of apparent neurological pathology recalling MS in these seropositive controls may indicate that they are healthy carriers and have eliminated an infectious virus after immunizing themselves, or that they constitute an at-risk population of chronic carriers. In effect, epidemiological data showing that a pathogenic agent present in the environment of regions of high prevalence of MS may be the cause of this disease imply that a fraction of the population free from MS has necessarily been in contact with such a pathogenic agent. It has been shown that the MSRV-1 retrovirus constitutes all or part of this "pathogenic agent" at the source of MS, and it is hence normal for controls taken from a healthy population to possess IgG type antibodies against components of the MSRV-1 retrovirus. Thus, the difference in seroprevalence between

the MS and control populations is extremely significant: "chi-squared" test,  $p < 0.001$ . These results hence point to an aetiopathogenic role of MSRV-1 in MS.

d) Detection of anti-MSRV-1 IgM antibodies by

5 ELISA:

The ELISA technique with the POL2B peptide was used to test for the presence of anti-MSRV-1 IgM specific antibodies in the serum of 36 patients for whom a definite or probable diagnosis of MS was established according to  
10 the criteria of Poser (23), and of 42 healthy controls (blood donors).

Figure 30 shows the results for each serum tested with an anti-IgM antibody. Each vertical bar represents the net optical density (OD at 492 nm) of a serum tested.  
15 The ordinate axis gives the net OD at the top of the vertical bars. The first 36 vertical bars lying to the left of the vertical line cutting the abscissa axis represent the sera of 36 cases of MS tested, and the vertical bars lying to the right of the vertical broken  
20 line represent the sera of 42 healthy controls (blood donors). The horizontal line drawn in the middle of the diagram represents a theoretical threshold defining the boundary of the positive results (in which the top of the bar lies above) and the negative results (in which the top  
25 of the bar lies below).

The mean of the net OD values for the MS cases tested is 0.19.

The mean of the net OD values for the controls is 0.09.

30 The standard deviation of the negative controls is 0.05.

In view of the small difference between the mean and the standard deviation of the controls, the threshold of theoretical positivity may be calculated according to  
35 the formula:

threshold value = (mean of the net OD values of the seronegative controls) + (3 x standard deviation of the net OD values of the seronegative controls).

5           The threshold value is hence equal to  $0.09 + (3 \times 0.05) = 0.26$ ; or, in practice, 0.25.

          The negative results represent a non-specific "background" of the presence of antibodies directed specifically against an epitope of the peptide.

10           According to this analysis, and as shown in Figure 30 and in the corresponding Table 2, the IgM test is specific for MS, since no control has a net OD above the threshold. 7 of the 36 MS sera produce a positive IgM result; now, a study of the clinical data reveals that  
15 these positive sera were taken during a first attack of MS or an acute attack in untreated patients. It is known that IgMs directed against pathogenic agents are produced during primary infections or during reactivations following a latency phase of the said pathogenic agent.

20           The difference in seroprevalence between the MS and control populations is extremely significant: "chi-squared" test,  $p < 0.001$ .

          These results point to an aetiopathogenic role of MSRV-1 in MS.

25           The detection of IgM and IgG antibodies against the POL2B peptide enables the course of an MSRV-1 infection and/or of the viral reactivation of MSRV-1 to be evaluated.

TABLE No. 2

	MS	CONTROLS
	0.064	0.243
	0.087	0.11
5	0.044	0.098
	0.115	0.028
	0.089	0.094
	0.025	0.038
	0.097	0.176
10	0.108	0.146
	0.018	0.049
	0.234	0.161
	0.274	0.113
	0.225	0.079
15	0.314	0.093
	0.522	0.127
	0.306	0.02
	0.143	0.052
	0.375	0.062
20	0.142	0.074
	0.157	0.043
	0.168	0.046
	1.051	0.041
	0.104	0.13
25	0.187	0.153
	0.044	0.107
	0.053	0.178
	0.153	0.114
	0.07	0.078
30	0.033	0.118
	0.104	0.177
	0.187	0.026
	0.044	0.024
	0.053	0.046
35	0.153	0.116
	0.07	0.04
	0.033	0.028
	0.973	0.073
		0.008
40		0.074
		0.141
		0.219
		0.047
		0.017
45	MEAN 0.19	0.09
	STD. DEV. 0.23	0.05
	THRESHOLD VALUE	0.26

e) Search for immunodominant epitopes in the POL2B peptide:

In order to reduce the non-specific background and to optimize the detection of the responses of the anti-MSRV-1 antibodies, the synthesis of octapeptides, advancing in successive one amino acid steps, covering the whole of the sequence determined by POL2B, was carried out according to the protocol described below.

The chemical synthesis of overlapping octapeptides covering the amino acid sequence 61-110 shown in the identifier SEQ ID NO:39 was carried out on an activated cellulose membrane according to the technique of BERG et al. (1989. J. Ann. Chem. Soc., 111, 8024-8026) marketed by Cambridge Research Biochemicals under the trade name Spotskan. This technique permits the simultaneous synthesis of a large number of peptides and their analysis.

The synthesis is carried out with esterified amino acids in which the  $\alpha$ -amino group is protected with an FMOC group (Nova Biochem) and the side-chain groups with protective groups such as trityl, t-butyl ester or t-butyl ether. The esterified amino acids are solubilized in N-methylpyrrolidone (NMP) at a concentration of 300 nM, and 0.9 ml are applied to spots of deposit of bromophenol blue. After incubation for 15 minutes, a further application of amino acids is carried out according to another 15-minute incubation. If the coupling between two amino acids has taken place correctly, a coloration modification (change from blue to yellow-green) is observed. After three washes in DMF, an acetylation step is performed with acetic anhydride. Next, the terminal amino groups of the peptides in the process of synthesis are deprotected with 20% pyridine in DMF. The spots of deposit are restained with a 1% solution of bromophenol blue in DMF, washed three times with methanol and dried. This set of operations constitutes one cycle of addition

of an amino acid, and this cycle is repeated until the synthesis is complete. When all the amino acids have been added, the NH<sub>2</sub>-terminal group of the last amino acid is deprotected with 20% piperidine in DMF and acetylated with  
5 acetic anhydride. The groups protecting the side chain are removed with a dichloromethane/trifluoroacetic acid/triisobutylsilane (5 ml/5 ml/250 ml) mixture. The immunoreactivity of the peptides is then tested by ELISA.

After synthesis of the different octapeptides in  
10 duplicate on two different membranes, the latter are rinsed with methanol and washed in TBS (0.1M Tris pH 7.2), then incubated overnight at room temperature in a saturation buffer. After several washes in TBS-T (0.1M Tris pH 7.2 - 0.05% Tween 20), one membrane is incubated  
15 with a 1/50 dilution of a reference serum originating from a patient suffering from MS, and the other membrane with a 1/50 dilution of a pool of sera of healthy controls. The membranes are incubated for 4 hours at room temperature. After washes with TBS-T, a  $\beta$ -galactosidase-labelled anti-  
20 human immunoglobulin conjugate (marketed by Cambridge Research Biochemicals) is added at a dilution of 1/200, and the mixture is incubated for two hours at room temperature. After washes of the membranes with 0.05% TBS-T and PBS, the immunoreactivity in the different spots is  
25 visualized by adding 5-bromo-4-chloro-3-indolyl  $\beta$ -D-galactopyranoside in potassium. The intensity of coloration of the spots is estimated qualitatively with a relative value from 0 to 5 as shown in the attached Figures 31 to 33.

30 In this way, it is possible to determine two immunodominant regions at each end of the POL2B peptide, corresponding, respectively, to the amino acid sequences 65-75 (SEQ ID NO:41) and 92-109 (SEQ ID NO:42), according to Figure 34, and lying, respectively, between the  
35 octapeptides Phe-Cys-Ile-Pro-Val-Arg-Pro-Asp (FCIPVRPD) and Arg-Pro-Asp-Ser-Gln-Phe-Leu-Phe (RPDSQFLF), and

Thr-Val-Leu-Pro-Gln-Gly-Phe-Arg (TVLPQGFR) and Leu-Phe-Gly-Gln-Ala-Leu-Ala-Gln (LFGQALAQ), and a region which is less reactive but apparently more specific, since it does not produce any background with the control serum,  
5 represented by the octapeptides Leu-Phe-Ala-Phe-Glu-Asp-Pro-Leu (LFAFEDPL) (SEQ ID NO:43) and Phe-Ala-Phe-Glu-Asp-Pro-Leu-Asn (FAFEDPLN) (SEQ ID NO:44).

These regions make it possible to define new peptides which are more specific and more immunoreactive  
10 according to the usual techniques.

It is thus possible, as a result of the discoveries made and the methods developed by the inventors, to carry out a diagnosis of MSRV-1 infection and/or reactivation and to evaluate a therapy in MS on the basis  
15 of its efficacy in "negativizing" the detection of these agents in the patients' biological fluids. Furthermore, early detection in individuals not yet displaying neurological signs of MS could make it possible to institute a treatment which would be all the more effective with  
20 respect to the subsequent clinical course for the fact that it would precede the lesion stage which corresponds to the onset of neurological disorders. Now, at the present time, a diagnosis of MS cannot be established before a symptomatology of neurological lesions has set  
25 in, and hence no treatment is instituted before the emergence of a clinical picture suggestive of lesions of the central nervous system which are already significant. The diagnosis of an MSRV-1 and/or MSRV-2 infection and/or reactivation in man is hence of decisive importance, and  
30 the present invention provides the means of doing this.

It is thus possible, apart from carrying out a diagnosis of MSRV-1 infection and/or reactivation, to evaluate a therapy in MS on the basis of its efficacy in "negativizing" the detection of these agents in the  
35 patients' biological fluids.



**EXAMPLE 12: OBTAINING A CLONE LB19 CONTAINING A PORTION OF THE gag GENE OF THE MSRV-1 RETROVIRUS**

A PCR technique derived from the technique published by Gonzalez-Quintial R et al. (19) and PLAZA et al. (25) was used. From the total RNAs extracted from a fraction of virion purified as described above, the cDNA was synthesized using a specific primer (SEQ ID No.64) at the 3' end of the genome to be amplified, using EXPAND™ REVERSE TRANSCRIPTASE (BOEHRINGER MANNHEIM).

10

cDNA:

AAGGGGCATG GACGAGGTGG TGGCTTATTT (SEQ ID NO:65)  
(antisense)

15

After purification, a poly(G) tail was added at the 5' end of the cDNA using the "Terminal transferases kit" marketed by the company Boehringer Mannheim, according to the manufacturer's protocol.

An anchoring PCR was carried out using the following 5' and 3' primers:

20 AGATCTGCAG AATTCGATAT CACCCCCCCC CCCCCC (SEQ ID No. 91)  
(sense), and AAATGTCTGC GGCACCAATC TCCATGTT  
(SEQ ID No. 64) (antisense)

Next, a semi-nested anchoring PCR was carried out with the following 5' and 3' primers:

25 AGATCTGCAG AATTCGATAT CA (SEQ ID No.92) (sense), and  
AAATGTCTGC GGCACCAATC TCCATGTT (SEQ ID No.64) (antisense)

The products originating from the PCR were purified after purification on agarose gel according to conventional methods (17), and then resuspended in 10 microlitres of distilled water. Since one of the properties of Taq polymerase consists in adding an adenine at the 3' end of each of the two DNA strands, the DNA obtained was inserted directly into a plasmid using the TA Cloning™ kit (British Biotechnology). The 2 µl of DNA solution were mixed with 5 µl of sterile distilled water,

30  
35

1  $\mu$ l of 10-fold concentrated ligation buffer "10x LIGATION  
BUFFER", 2  $\mu$ l of "pCR<sup>TM</sup> VECTOR" (25 ng/ml) and 1  $\mu$ l of "T4  
DNA LIGASE". This mixture was incubated overnight at 12°C.  
The following steps were carried out according to the  
5 instructions of the TA Cloning<sup>TM</sup> kit (British  
Biotechnology). At the end of the procedure, the white  
colonies of recombinant bacteria (white) were picked out  
in order to be cultured and to permit extraction of the  
plasmids incorporated according to the so-called  
10 "miniprep" procedure (17). The plasmid preparation from  
each recombinant colony was cut with a suitable  
restriction enzyme and analysed on agarose gel. Plasmids  
possessing an insert detected under UV light after  
staining the gel with ethidium bromide were selected for  
15 sequencing of the insert, after hybridization with a  
primer complementary to the Sp6 promoter present on the  
cloning plasmid of the TA Cloning Kit<sup>TM</sup>. The reaction prior  
to sequencing was then performed according to the method  
recommended for the use of the sequencing kit "Prism ready  
20 reaction kit dye deoxyterminator cycle sequencing kit"  
(Applied Biosystems, ref. 401384), and automatic  
sequencing was carried out with an Applied Biosystems  
"Automatic Sequencer, model 373 A" apparatus according to  
the manufacturer's instructions.  
25 PCR amplification according to the technique  
mentioned above was used on a cDNA synthesized from the  
nucleic acids of fractions of infective particles purified  
on a sucrose gradient, according to the technique  
described by H. Perron (13), from culture supernatants of  
30 B lymphocytes of a patient suffering from MS, immortalized  
with Epstein-Barr virus (EBV) strain B95 and expressing  
retroviral particles associated with reverse transcriptase  
activity as described by Perron et al. (3) and in French  
Patent Applications MS 10, 11 and 12. the clone LB19,  
35 whose sequence, identified by SEQ ID NO:59, is presented  
in Figure 35.

The clone makes it possible to define, with the clone GM3 previously sequenced and the clone G+E+A (see Example 15), a region of 690 base pairs representative of a significant portion of the gag gene of the MSRV-1 retrovirus, as presented in Figure 36. This sequence designated SEQ ID NO:88 is reconstituted from different clones overlapping at their ends. This sequence is identified under the name MSRV-1 "gag\*" region. In Figure 36, a potential reading frame with the translation into amino acids is presented below the nucleic acid sequence.

**EXAMPLE 13: OBTAINING A CLONE FBd13 CONTAINING A pol GENE REGION RELATED TO THE MSRV-1 RETROVIRUS AND AN APPARENTLY INCOMPLETE ENV REGION CONTAINING A POTENTIAL READING FRAME (ORF) FOR A GLYCOPROTEIN**

Extraction of viral RNAs: The RNAs were extracted according to the method briefly described below.

A pool of culture supernatant of B lymphocytes of patients suffering from MS (650 ml) is centrifuged for 30 minutes at 10,000 g. The viral pellet obtained is resuspended in 300 microlitres of PBS/10 mM MgCl<sub>2</sub>. The material is treated with a DNase (100 mg/ml)/RNase (50 mg/ml) mixture for 30 minutes at 37°C and then with proteinase K (50 mg/ml) for 30 minutes at 46°C.

The nucleic acids are extracted with one volume of a phenol/0.1% SDS (V/V) mixture heated to 60°C, and then re-extracted with one volume of phenol/chloroform (1:1; V/V).

Precipitation of the material is performed with 2.5 V of ethanol in the presence of 0.1 V of sodium acetate pH5.2. The pellet obtained after centrifugation is resuspended in 50 microlitres of sterile DEPC water.

The sample is treated again with 50 mg/ml of "RNase free" DNase for 30 minutes at room temperature, extracted with one volume of phenol/chloroform and

precipitated in the presence of sodium acetate and ethanol.

The RNA obtained is quantified by an OD reading at 260 nm. The presence of MSRV-1 and the absence of DNA  
5 contaminant is monitored by a PCR and an MSRV-1-specific RTPCR associated with a specific ELOSA for the MSRV-1 genome.

#### Synthesis of cDNA:

5 mg of RNA are used to synthesize a cDNA primed  
10 with a poly(DT) oligonucleotide according to the instructions of the "cDNA Synthesis Module" kit (ref RPN 1256, Amersham) with a few modifications: The reverse transcription is performed at 45°C instead of the recommended 42°C.

15 The synthesis product is purified by a double extraction and a double purification according to the manufacturer's instructions.

The presence of MSRV-1 is verified by an MSRV-1 PCR associated with a specific ELOSA for the MSRV-1  
20 genome.

#### "Long Distance PCR": (LD-PCR)

500 ng of cDNA are used for the LD-PCR step (Expand Long Template System; Boehringer (ref.1681 842)).

Several pairs of oligonucleotides were used.

25 Among these, the pair defined by the following primers:

5' primer: GGAGAAGAGC AGCATAAGTG G (SEQ ID NO:66)

3' primer: GTGCTGATTG GTGTATTTAC AATCC (SEQ ID NO:67).

The amplification conditions are as follows:

94°C 10 seconds  
30 56°C 30 seconds  
68°C 5 minutes;

10 cycles, then 20 cycles with an increment of 20 seconds in each cycle on the elongation time. At the end of this first amplification, 2 microlitres of the  
35 amplification product are subjected to a second amplification under the same conditions as before.

The LD-PCR reactions are conducted in a Perkin model 9600 PCR apparatus in thin-walled microtubes (Boehringer).

5 The amplification products are monitored by electrophoresis of 1/5th of the amplification volume (10 microlitres) in 1% agarose gel. For the pair of primers described above, a band of approximately 1.7 Kb is obtained.

Cloning of the amplified fragment:

10 The PCR product was purified by passage through a preparative agarose gel and then through a Costar column (Spin; D. Dutcher) according to the supplier's instructions.

15 2 microlitres of the purified solution are joined up with 50 ng of vector PCRII according to the supplier's instructions (TA Cloning Kit; British Biotechnology)).

20 The recombinant vector obtained is isolated by transformation of competent DH5 $\alpha$ F' bacteria. The bacteria are selected using their resistance to ampicillin and the loss of metabolism for Xgal (= white colonies). The molecular structure of the recombinant vector is confirmed by plasmid minipreparation and hydrolysis with the enzyme EcoRI.

25 FBd13, a positive clone for all these criteria, was selected. A large-scale preparation of the recombinant plasmid was performed using the Midiprep Quiagen kit (ref 12243) according to the supplier's instructions.

30 Sequencing of the clone FBd13 is performed by means of the Perkin Prism Ready Amplitaq FS dye terminator kit (ref. 402119) according to the manufacturer's instructions. The sequence reactions are introduced into a Perkin type 377 or 373A automatic sequencer. The sequencing strategy consists in gene walking carried out  
35 on both strands of the clone Fbd13.

The sequence of the clone FBd13 is identified by SEQ ID NO:58.

In Figure 37, the sequence homology between the clone FBd13 and the HSERV-9 retrovirus is shown on the matrix chart by a continuous line for any partial homology greater than or equal to 70%. It can be seen that there are homologies in the flanking regions of the clone (with the pol gene at the 5' end and with the env gene and then the LTR at the 3' end), but that the internal region is totally divergent and does not display any homology, even weak, with the env gene of HSERV-9. Furthermore, it is apparent that the clone FBd13 contains a longer "env" region than the one which is described for the defective endogenous HSERV-9; it may thus be seen that the internal divergent region constitutes an "insert" between the regions of partial homology with the HSERV-9 defective genes.

This additional sequence determines a potential orf, designated ORF B13, which is represented by its amino acid sequence SEQ ID NO:87.

The molecular structure of the clone FBd13 was analyzed using the GeneWork software and Genbank and SwissProt data banks.

5 glycosylation sites were found.

The protein does not have significant homology with already known sequences.

It is probable that this clone originates from a recombination of an endogenous retroviral element (ERV), linked to the replication of MSRV-1.

Such a phenomenon does not lack generation of the expression of polypeptides, or even of endogenous retroviral proteins which are not necessarily tolerated by the immune system. Such a scheme of aberrant expression of endogenous elements related to MSRV-1 and/or induced by the latter is liable to multiply the aberrant antigens, and hence tends to contribute to the induction of

autoimmune processes such as are observed in MS. It clearly constitutes a novel element never hitherto described. In effect, interrogation of the data banks of nucleic acid sequences available in version No. 19 (1996) of the "Entrez" software (NCBI, NIH, Bethesda, USA) did not enable a known homologous sequence comprising the whole of the env region of this clone to be identified.

EXAMPLE 14: OBTAINING A CLONE FP6 CONTAINING A PORTION OF THE pol GENE, WITH A REGION CODING FOR THE REVERSE TRANSCRIPTASE ENZYME HOMOLOGOUS TO THE CLONE POL\*MSRV-1, AND A 3'pol REGION DIVERGENT FROM THE EQUIVALENT SEQUENCES DESCRIBED IN THE CLONES POL\*, tpol, FBd3, JLBc1 and JLBc2

A 3'RACE was performed on total RNA extracted from plasma of a patient suffering from MS. A healthy control plasma treated under the same conditions was used as negative control. The synthesis of cDNA was carried out with the following modified oligo(dT) primer:

5' GACTCGCTGC AGATCGATTT TTTTTTTTTT TTTT 3' (SEQ ID NO:68)  
and Boehringer "Expand RT" reverse transcriptase according to the conditions recommended by the company. A PCR was performed with the enzyme Klentaq (Clontech) under the following conditions: 94°C 5 min then 93°C 1 min, 58°C 1 min, 68°C 3 min for 40 cycles and 68°C for 8 min, and with a final reaction volume of 50 µl.

Primers used for the PCR:

- 5' primer, identified by SEQ ID NO:69  
5' GCCATCAAGC CACCCAAGAA CTCTTAACTT 3';  
- 3' primer, identified by SEQ ID NO:68 (=the same as for the cDNA)

A second, so-called "semi-nested" PCR was carried out with a 5' primer located within the region already amplified. This second PCR was performed under the same experimental conditions as those used in the first

PCR, using 10  $\mu$ l of the amplification product originating from the first PCR.

Primers used for the semi-nested PCR:

- 5' primer, identified by SEQ ID NO:70

5' CCAATAGCCA GACCATTATA TACACTAATT 3';

- 3' primer, identified by SEQ ID NO:68 (=the same as for the cDNA)

Primers SEQ ID NO:69 and SEQ ID NO:70 are specific for the pol\* region: position No. 403 to No. 422 and No. 641 to No. 670, respectively.

An amplification product was thus obtained from the extracellular RNA extracted from the plasma of a patient suffering from MS. The corresponding fragment was not observed for the plasma of the healthy control. This amplification product was cloned in the following manner.

The amplified DNA was inserted into a plasmid using the TA Cloning™ kit. The 2  $\mu$ l of DNA solution were mixed with 5  $\mu$ l of sterile distilled water, 1  $\mu$ l of a 10-fold concentrated ligation buffer "10x LIGATION BUFFER", 2  $\mu$ l of "pCR™ VECTOR" (25 ng/ml) and 1  $\mu$ l of "TA DNA LIGASE". This mixture was incubated overnight at 12°C. The following steps were carried out according to the instructions of the TA Cloning™ kit (British Biotechnology). At the end of the procedure, the white columns of recombinant bacteria (white) were picked out in order to be cultured and to permit extraction of the plasmids incorporated according to the so-called "miniprep" procedure (17). The plasmid preparation from each recombinant colony was cut with a suitable restriction enzyme and analyzed on agarose gel. Plasmids possessing an insert detected under UV light after staining the gel with ethidium bromide was selected for sequencing of the insert, after hybridization with a primer complementary to the Sp6 promoter present on the cloning plasmid of the TA cloning kit™. The reaction prior to sequencing was then performed according to the method



recommended for the use of the sequencing kit "Prism ready reaction kit dye deoxyterminator cycle sequencing kit" (Applied Biosystems, ref. 401384), and automatic sequencing was carried out with an Applied Biosystems  
5 "Automatic Sequencer, model 373 A" apparatus according to the manufacturer's instructions.

The clone obtained, designated FP6, enables a region of 467 bp which is 89% homologous to the pol\* region of the MSRV-1 retrovirus and a region of 1167 bp  
10 which is 64% homologous to the pol region of ERV-9 (No. 1634 to 2856) to be defined.

The clone FP6 is represented in Figure 38 by its nucleotide sequence identified by SEQ ID NO:61. The three potential reading frames of this clone are indicated by  
15 their amino acid sequence under the nucleotide sequence.

**EXAMPLE 15: OBTAINING A REGION DESIGNATED G+E+A CONTAINING AN ORF FOR A RETROVIRAL PROTEASE, BY PCR AMPLIFICATION OF THE NUCLEIC ACID SEQUENCE CONTAINED  
20 BETWEEN THE 5' REGION DEFINED BY THE CLONE "GM3" AND THE 3' REGION DEFINED BY THE CLONE POL\*, FROM THE RNA EXTRACTED FROM A POOL OF PLASMAS OF PATIENTS SUFFERING FROM MS**

Oligonucleotides specific for the MSRV-1  
25 sequences already identified by the Applicant were defined in order to amplify the retroviral RNA originating from virions present in the plasma of patients suffering from MS. Control reactions were performed so as to monitor the presence of contaminants (reaction with water). The  
30 amplification consists of a step of RT-PCR followed by a "nested" PCR. Pairs of primers were defined for amplifying three overlapping regions (designated G, E and A) on the regions defined by the sequences of the clones GM3 and pol\* described above.

35

Semi-nested RT-PCR for amplification of the region G:

- in the first RT-PCR cycle, the following primers are used:

primer 1: SEQ ID NO:71 (sense)

primer 2: SEQ ID NO:72 (antisense)

5 - in the second PCR cycle, the following primers are used:

primer 1: SEQ ID NO:73 (sense)

primer 4: SEQ ID NO:74 (antisense)

Nested RT-PCR for amplification of the region E:

10 - in the first RT-PCR cycle, the following primers are used:

primer 5: SEQ ID NO:75 (sense)

primer 6: SEQ ID NO:76 (antisense)

15 - in the second PCR cycle, the following primers are used:

primer 7: SEQ ID NO:77 (sense)

primer 8: SEQ ID NO:78 (antisense)

Semi-nested RT-PCR for amplification of the region A:

20 - in the first RT-PCR cycle, the following primers are used:

primer 9: SEQ ID NO:79 (sense)

primer 10: SEQ ID NO:80 (antisense)

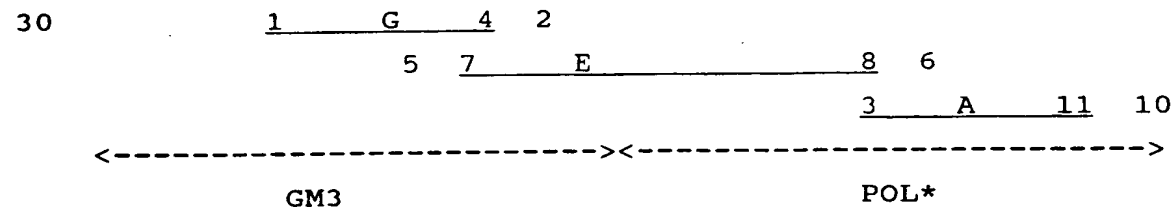
- in the second PCR cycle, the following primers are used:

25 primer 9: SEQ ID NO:81 (sense)

primer 11: SEQ ID NO:82 (antisense)

The primers and the regions G, E and A which they define are positioned as follows:

cDNA



The sequence of the region defined by the different clones G, E and A was determined after cloning and sequencing of the "nested" amplification products.

The clones G, E and A were assembled together by  
5 PCR with the primers 1 at the 5' end of the fragment G and  
11 at the 3' end of the fragment A, the primers being  
described above. An approximately 1580-bp fragment G+E+A  
was amplified and inserted into a plasmid using the TA  
Cloning (trademark) kit. The sequence of the amplification  
10 product corresponding to G+E+A was determined and analysis  
of the G+E and E+A overlaps was carried out. The sequence  
is shown in Figure 39, and corresponds to the sequence SEQ  
ID NO:89.

A reading frame coding for an MSRV-1 retroviral  
15 protease was found in the region E. The amino acid  
sequence of the protease, identified by SEQ ID NO:90, is  
presented in Figure 40.

**EXAMPLE 16: OBTAINING A CLONE LTRGAG12, RELATED  
20 TO AN ENDOGENOUS RETROVIRAL ELEMENT (ERV) CLOSE TO MSRV-1,  
IN THE DNA OF AN MS LYMPHOBLASTOID LINE PRODUCING VIRIONS  
AND EXPRESSING THE MSRV-1 RETROVIRUS**

A nested PCR was performed on the DNA extracted  
from a lymphoblastoid line (B lymphocytes immortalized  
25 with the EBV virus strain B95, as described above and as  
is well known to a person skilled in the art) expressing  
the MSRV-1 retrovirus and originating from peripheral  
blood lymphocytes of a patient suffering from MS.

In the first PCR step, the following primers are  
30 used:

primer 4327: CTCGATTTCT TGCTGGGCCT TA (SEQ ID NO:83)

primer 3512: GTTGATTCCC TCCTCAAGCA (SEQ ID NO:84)

This step comprises 35 amplification cycles with  
the following conditions: 1 min at 94°C, 1 min at 54°C and  
35 4 min at 72°C.

In the second PCR step, the following primers are used:

primer 4294: CTCTACCAAT CAGCATGTGG (SEQ ID NO:85)

primer 3591: TGTTCTCTT GGTCCTAT (SEQ ID NO:86)

- 5 This step comprises 35 amplification cycles with the following conditions: 1 min at 94°C, 1 min at 54°C and 4 min at 72°C.

The products originating from the PCR were purified after purification on agarose gel according to  
10 conventional methods (17), and then resuspended in 10 ml of distilled water. Since one of the properties of Taq polymerase consists in adding an adenine at the 3' end of each of the two DNA strands, the DNA obtained was inserted directly into a plasmid using the TA Cloning™ kit (British  
15 Biotechnology). The 2 µl of DNA solution were mixed with 5 µl of sterile distilled water, 1 µl of a 10-fold concentrated ligation buffer "10x LIGATION BUFFER", 2 µl of "pCR™ VECTOR" (25 ng/ml) and 1 µl of "TA DNA LIGASE". This mixture was incubated overnight at 12°C. The  
20 following steps were carried out according to the instructions of the TA Cloning™ kit (British Biotechnology). At the end of the procedure, the white colonies of recombinant bacteria (white) were picked out in order to be cultured and to permit extraction of the  
25 plasmids incorporated according to the so-called "miniprep" procedure (17). The plasmid preparation from each recombinant colony was cut with a suitable restriction enzyme and analyzed on agarose gel. The plasmids possessing an insert detected under UV light  
30 after staining the gel with ethidium bromide were selected for sequencing of the insert, after hybridization with a primer complementary to the Sp6 promoter present on the cloning plasmid of the TA Cloning Kit™. The reaction prior to sequencing was then performed according to the method  
35 recommended for the use of the sequencing kit "Prism ready reaction kit dye deoxyterminator cycle sequencing kit"

(Applied Biosystems, ref. 401384), and automatic sequencing was carried out with an Applied Biosystems "Automatic Sequencer, model 373 A" apparatus according to the manufacturer's instructions.

5           Thus, a clone designated LTRGAG12 could be  
obtained, and is represented by its internal sequence  
identified by SEQ ID NO:60.

This clone is probably representative of endogenous elements close to ERV-9, present in human DNA, in particular in the DNA of patients suffering from MS, and capable of interfering with the expression of the MSRV-1 retrovirus, hence capable of having a role in the pathogenesis associated with the MSRV-1 retrovirus and capable of serving as marker for a specific expression in the pathology in question.

EXAMPLE 17: DETECTION OF ANTI-MSRV-1 SPECIFIC ANTIBODIES IN HUMAN SERUM

Identification of the sequence of the pol gene  
20 of the MSRV-1 retrovirus and of an open reading frame of  
this gene enabled the amino acid sequence SEQ ID NO:63 of  
a region of the said gene, referenced SEQ ID NO:62, to be  
determined.

Different synthetic peptides corresponding to  
25 fragments of the protein sequence of MSRV-1 reverse  
transcriptase encoded by the pol gene were tested for  
their antigenic specificity with respect to sera of  
patients suffering from MS and of healthy controls.

The peptides were synthesized chemically by solid-phase synthesis according to the Merrifield technique (22). The practical details are those described below.

a) Peptide synthesis:

The peptides were synthesized on a phenylacet-  
amidomethyl (PAM)/polystyrene/divinylbenzene resin  
(Applied Biosystems, Inc. Foster City, CA), using an

"Applied Biosystems 430A" automatic synthesizer. The amino acids are coupled in the form of hydroxybenzotriazole (HOBT) esters. The amino acids used are obtained from Novabiochem (Läufelfingen, Switzerland) or Bachem (Bubendorf, Switzerland).

The chemical synthesis was performed using a double coupling protocol with N-methylpyrrolidone (NMP) as solvent. The peptides were cut from the resin, as well as the side-chain protective groups, simultaneously, using hydrofluoric acid (HF) in a suitable apparatus (type I cleavage apparatus, Peptide Institute, Osaka, Japan).

For 1 g of peptidyl resin, 10 ml of HF, 1 ml of anisole and 1 ml of dimethyl sulphide (DMS) are used. The mixture is stirred for 45 minutes at -2°C. The HF is then evaporated off under vacuum. After intensive washes with ether, the peptide is eluted from the resin with 10% acetic acid and then lyophilized.

The peptides are purified by preparative high performance liquid chromatography on a VYDAC C18 type column (250 x 21 mm) (The Separation Group, Hesperia, CA, USA). Elution is carried out with an acetonitrile gradient at a flow rate of 22 ml/min. The fractions collected are monitored by an elution under isocratic conditions on a VYDAC<sup>TM</sup> C18 analytical column (250 x 4.6 mm) at a flow rate of 1 ml/min. Fractions having the same retention time are pooled and lyophilized. The preponderant fraction is then analysed by analytical high performance liquid chromatography with the system described above. The peptide which is considered to be of acceptable purity manifests itself in a single peak representing not less than 95% of the chromatogram.

The purified peptides are then analysed with the object of monitoring their amino acid composition, using an Applied Biosystems 420H automatic amino acid analyser. Measurement of the (average) chemical molecular mass of the peptides is obtained using LSIMS mass spectrometry in

the positive ion mode on a VG. ZAB.ZSEQ double focusing instrument connected to a DEC-VAX 2000 acquisition system (VG analytical Ltd, Manchester, England).

5 The reactivity of the different peptides was tested against sera of patients suffering from MS and against sera of healthy controls. This enabled a peptide designated S24Q to be selected, whose sequence is identified by SEQ ID NO:63, encoded by a nucleotide sequence of the pol gene of MSRV-1 (SEQ ID NO:62).

10

b) Antigenic properties:

The antigenic properties of the S24Q peptide were demonstrated according to the ELISA protocol described below.

15

The lyophilized S24Q peptide was dissolved in 10 % acetic acid at a concentration of 1 mg/ml. This stock solution was aliquoted and kept at +4°C for use over a fortnight, or frozen at -20°C for use within 2 months. An aliquot is diluted in PBS (phosphate buffered saline) solution so as to obtain a final peptide concentration of 5 micrograms/ml. 100 microlitres of this dilution are placed in each well of Nunc Maxisorb (trade name) microtitration plates. The plates are covered with a "plate-sealer" type adhesive and kept for 2 hours at +37°C

20 for the phase of adsorption of the peptide to the plastic. The adhesive is removed and the plates are washed three times with a volume of 300 microlitres of a solution A (1X PBS, 0.05% Tween 20®), then inverted over an absorbent tissue. The plates thus drained are filled with

25 250 microlitres per well of a solution B (solution A + 10% of goat serum), then covered with an adhesive and incubated for 1 hour at 37°C. The plates are then washed three times with the solution A as described above.

30 The test serum samples are diluted beforehand to 1/100 in the solution B, and 100 microlitres of each dilute test serum are placed in the wells of each micro-

titration plate. A negative control is placed in one well of each plate, in the form of 100 microlitres of buffer B. The plates covered with an adhesive are then incubated for 1 hour 30 min at 37°C. The plates are then washed three  
5 times with the solution A as described above. For the IgG response, a peroxidase-labelled goat antibody directed against human IgG (marketed by Jackson Immuno Research Inc.) is diluted in the solution B (dilution 1/10,000). 100 microlitres of the appropriate dilution of the  
10 labelled antibody are then placed in each well of the microtitration plates, and the plates covered with an adhesive are incubated for 1 hour at 37°C. A further washing of the plates is then performed as described above. In parallel, the peroxidase substrate is prepared  
15 according to the directions of the bioMérieux kits. 100 microlitres of substrate solution are placed in each well, and the plates are placed protected from light for 20 to 30 minutes at room temperature.

When the colour reaction has stabilized,  
20 50 microlitres of Color 2 (bioMérieux trade name) are placed in each well in order to stop the reaction. The plates are placed immediately in an ELISA plate spectrophotometric reader, and the optical density (OD) of each well is read at a wavelength of 492 nm.

25 The serological samples are introduced in duplicate or in triplicate, and the optical density (OD) corresponding to the serum tested is calculated by taking the mean of the OD values obtained for the same sample at the same dilution.

30 The net OD of each serum corresponds to the mean OD of the serum minus the mean OD of the negative control (solution B: PBS, 0.05% Tween 20x, 10% goat serum).

c) Detection of anti-MSRV-1 IgG antibodies (S24Q) by ELISA:

35 The technique described above was used with the S24Q peptide to test for the presence of anti-MSRV-1



specific IgG antibodies in the serum of 15 patients for whom a definite diagnosis of MS was established according to the criteria of Poser (23), and of 15 healthy controls (blood donors).

5           Figure 41 shows the results for each serum tested with an anti-IgG antibody. Each vertical bar represents the net optical density (OD at 492 nm) of a serum tested. The ordinate axis gives the net OD at the top of the vertical bars. The first 15 vertical bars lying  
10 to the left of the vertical broken line represent the sera of 15 healthy controls (blood donors), and the 15 vertical bars lying to the right of the vertical broken line represent the sera of 15 cases of MS tested. The diagram enables 2 controls to be revealed whose OD rises above the  
15 grouped values of the control population. These values may represent the presence of specific IgGs in symptomless seropositive patients. Two methods were hence evaluated in order to determine the statistical threshold of positivity of the test.

20           The mean of the net OD values for the controls, including the controls with high net OD values, is 0.129 and the standard deviation is 0.06. Without the 2 controls whose OD values are greater than 0.2, the mean of the "negative" controls is 0.107 and the standard deviation is  
25 0.03. A theoretical threshold of positivity may be calculated according to the formula:

threshold value (mean of the net OD values of the negative controls) + ( 2 or 3  $\sqrt$  standard deviation  
30 of the net OD values of the negative controls).

In the first case, there are considered to be symptomless seropositives, and the threshold value is equal to  $0.11 + (3 \times 0.03) = 0.20$ . The negative results  
35 represent a non-specific "background" of the presence of

antibodies directed specifically against an epitope of the peptide.

In the second case, if the set of controls consisting of blood donors in apparent good health is taken as a reference basis, without excluding the sera which are, on the face of it, seropositive, the standard deviation of the "non-MS controls" is 0.116. The threshold value then becomes  $0.13 + (3 \times 0.06) = 0.31$ .

According to this latter analysis, the test is specific for MS. In this respect, it is seen that the test is specific for MS, since, as shown in Table 1, no control has a net OD above this threshold. In fact, this result reflects the fact that the antibody titres in patients suffering from MS are, for the most part, higher than in healthy controls who have been in contact with MSRV-1.

In accordance with the first method of calculation, and as shown in Figure 41 and in Table 3, 6 of the 15 MS sera give a positive result (OD greater than or equal to 0.2), indicating the presence of IgGs specifically directed against the S24Q peptide, hence against a portion of the reverse transcriptase enzyme of the MSRV-1 retrovirus encoded by its pol gene, and consequently against the MSRV-1 retrovirus.

Thus, approximately 40% of the MS patients tested have reacted against an epitope carried by the S24Q peptide and possess circulating IgGs directed against the latter.

Two out of 15 blood donors in apparent good health show a positive result. Thus, it is apparent that approximately 13% of the symptomless population may have been in contact with an epitope carried by the S24Q peptide under conditions which have led to an active immunization which manifests itself in the persistence of specific serum IgGs. These conditions are compatible with an immunization against the MSRV-1 retrovirus reverse transcriptase during an infection with (and/or reactiva-

tion of) the MSRV-1 retrovirus. The absence of apparent neurological pathology recalling MS in these seropositive controls may indicate that they are healthy carriers and have eliminated an infectious virus after immunizing  
5 themselves, or that they constitute an at-risk population of chronic carriers. In effect, epidemiological data showing that a pathogenic agent present in the environment of regions of high prevalence of MS may be the cause of this disease imply that a fraction of the population free  
10 from MS has necessarily been in contact with such a pathogenic agent. It has been shown that the MSRV-1 retrovirus constitutes all or part of this "pathogenic agent" at the source of MS, and it is hence normal for controls taken from a healthy population to possess IgG  
15 type antibodies against components of the MSRV-1 retrovirus.

Lastly, the detection of anti-S24Q antibodies in only one out of two MS cases tested here may reflect the fact that this peptide does not represent an  
20 immunodominant MSRV-1 epitope, that inter-individual strain variations may induce an immunization against a divergent peptide motif in the same region, or that the course of the disease and the treatments followed may modulate over time the antibody response against the S24Q  
25 peptide.

30

35

TABLE No. 3

	CONTROLS	MS
	0.101	0.136
	0.058	0.391
5	0.126	0.37
	0.131	0.119
	0.105	0.267
	0.294	0.141
	0.116	0.102
	0.088	0.18
	0.105	0.411
	0.172	0.164
10	0.137	0.049
	0.223	0.644
	0.08	0.268
	0.073	0.065
	0.132	0.074
	Mean	0.129
	Std. Dev.	0.06
15	Threshold	0.31

d) Detection of anti-MSRV-1 IgM antibodies by ELISA:

20 The ELISA technique with the S24Q peptide was used to test for the presence of anti-MSRV-1 IgM specific antibodies in the same sera as above.

Figure 42 shows the results for each serum tested with an anti-IgM antibody. Each vertical bar represents the net optical density (OD at 492 nm) of a serum tested. The ordinate axis gives the net OD at the top of the vertical bars. The first 15 vertical bars lying to the left of the vertical line cutting the abscissa axis represent the sera of 15 healthy controls (blood donors), and the vertical bars lying to the right of the vertical broken line represent the sera of 15 cases of MS tested.

The mean of the OD values for the MS cases tested is 1.6.

The mean of the net OD values for the controls is 0.7.

The standard deviation of the negative controls is 0.6.

The threshold of theoretical positivity may be calculated according to the formula:

5

threshold value = (mean of the OD values of the negative controls) + (3 x standard deviation of the OD values of the negative controls)

10 The threshold value is hence equal to  $0.7 + (3 \times 0.6) = 2.5$ ;

The negative results represent a non-specific "background" of the presence of antibodies directed specifically against an epitope of the peptide.

15

According to this analysis, and as shown in Figure 42 and in the corresponding Table 4, the IgM test is specific for MS, since no control has a net OD above the threshold. 6 of the 15 MS sera produce a positive IgM result

20

The difference in seroprevalence between the MS and control populations is extremely significant: "chi-squared" test,  $p < 0.002$ .

These results point to an aetiopathogenic role of MSRV-1 in MS.

25

Thus, the detection of IgM and IgG antibodies against the S24Q peptide makes it possible to evaluate, alone or in combination with other MSRV-1 peptides, the course of an MSRV-1 infection and/or of the viral reactivation of MSRV-1.

TABLE No. 4

	CONTROLS	MS
	1.449	0.974
	0.371	6.117
	0.448	2.883
5	0.456	1.945
	0.885	1.787
	2.235	0.273
	0.301	1.766
	0.138	0.668
	0.16	2.603
	1.073	0.802
10	1.366	0.245
	0.283	0.147
	0.262	2.441
	0.585	0.287
	0.356	0.589
	Mean	0.7
	Std. Dev.	0.6
15	Threshold	
	Value	2.5

It is possible, as a result of the new discoveries made and the new methods developed by the inventors, to permit the improved implementation of diagnostic tests for MSRV-1 infection and/or reactivation and to evaluate a therapy in MS and/or RA on the basis of its efficacy in "negativizing" the detection of these agents in the patient's biological fluids. Furthermore, early detection in individuals not yet displaying neurological signs of MS or rheumatological signs of RA could make it possible to institute a treatment which would be all the more effective with respect to the subsequent clinical course for the fact that it would precede the lesion stage which corresponds to the onset of the clinical disorders. Now, at the present time, a diagnosis of MS or RA cannot be established before a symptomatology of lesions has set in, and hence no treatment is instituted before the emergence of a clinical picture suggestive of lesions which are already significant. The diagnosis of an MSRV-1 and/or MSRV-2 infection and/or reactivation in man is

hence of decisive importance, and the present invention provides the means of doing this.

It is thus possible, apart from carrying out a diagnosis of MSRV-1 infection and/or reactivation, to  
5 evaluate a therapy in MS on the basis of its efficacy in "negativizing" the detection of these agents in the patients' biological fluids.

**EXAMPLE 18 :**

10 1) MATERIALS AND METHODS

- Patients and clinical samples

Choroid plexus cells from MS patients and controls were obtained from the brain-cell library, Laboratoire R. Escourolles, Hôpital de la Salpêtrière,  
15 Paris, France. Non-tumoral leptomeningeal cells from controls were obtained as previously described (26). Peripheral blood from MS and control patients used for obtaining B-cell lines and plasma, were obtained from the Neurological Departments, CHU de Grenoble, and from  
20 INSERM U 134, Hôpital de la Salpêtrière, France. Clinical details and origin of the 10 MS patients and of the 10 patients with other neurological diseases who provided CSF samples are given in Table 6.

- Cell cultures, virus isolation and purification

25 All cell-types were cultured as previously described (3, 5, 26).

All cultures were regularly screened for mycoplasma contamination with an ELISA mycoplasma-detection kit (Boehringer). No cell-extract nor supernatant used  
30 contained detectable mycoplasma.

Extracellular virion purification and sucrose density gradients were performed as previously described (3, 5, 26). From each sucrose gradient 0.5-1ml fractions were collected from the top of the tubes, with a 1000µl  
35 Pipetman and a different sterile tip for each fraction. 60µl were used for RT activity assay and the rest was

mixed with 1 volume of buffer containing 4M guanidinium thiocyanate, 0.5% N-Lauroyl sarcosine, 25mM EDTA, 0.2%  $\beta$ -mercaptoethanol adjusted at pH 5.5 with acetic acid. These mixtures were frozen at  $-80^{\circ}\text{C}$  for further RNA extraction or directly processed according to Chomzynski (20), with an overnight precipitation step at  $-20^{\circ}\text{C}$ , in presence of RNase-free glycogen (Boehringer). RNA was dissolved 20 to 50  $\mu\text{l}$  of DEPC-treated water in the presence of 1-2  $\mu\text{l}$  of recombinant RNase-inhibitor (PROMEGA) and 0,1mM DTT. 10  $\mu\text{l}$  aliquots were used for each RT-PCR.

- Reverse transcriptase activity

RT-activity was tested with 20mM  $\text{Mg}^{++}$  and poly-Cm or polyC templates, in virion pellets or fractions from sucrose gradients as previously described (3, 5, 26).

- cDNA synthesis and 'Pan-retro' RT-PCR with degenerate primers

A total RT-activity between  $10^6$ - $10^7$  dpm was required in the fraction containing the peak of purified virions. The "Pan-retro" RT-PCR technique (27) was performed on virion RNA extracted by the method of Chomczynski (20) and dissolved in 20  $\mu\text{l}$  RNase-free water. 5  $\mu\text{l}$  RNA solution was incubated for 30 min at  $37^{\circ}\text{C}$  with 0.3 units (3 units for CSF series) of RNase-free DNase-1 (Boehringer) in a 20  $\mu\text{l}$  reaction containing 7.5 mM random hexamers, 5 mM Hepes-HCl pH 6.9, 75 mM KCl, 3 mM  $\text{MgCl}_2$ , 10 mM DTT, 50 mM Tris-HCl pH 7.5, 0.5 mM each dNTP, and 20 units recombinant RNase inhibitor (Promega). The DNase was then heat inactivated at  $80^{\circ}\text{C}$  for 10 min. 20 units MoMLV RT (Pharmacia) and a further 20 units of RNase inhibitor were added to each tube in a Genesphere<sup>TM</sup> enclosure (Safetech, Ireland) and cDNA was synthesised for 90 min at  $37^{\circ}\text{C}$ . Following reverse transcription, the cDNA was boiled for 5 min then cooled rapidly on ice. The Round 1 PCR mix (final volume 25  $\mu\text{l}$  per reaction; 20 mM Tris-HCl pH 8.4, 60 mM KCl, 2.5 mM  $\text{MgCl}_2$ , 200 ng each of primers PAN-UO and PAN-DI [see Figure 44], 0.2 mM each dNTP) was treated with



0.3 units DNase-1 and then heat inactivated as above. 2.5  $\mu$ l cDNA was added in the Genesphere<sup>TM</sup> enclosure and the tubes heated to 80°C before adding 0.5 units Taq polymerase (Perkin Elmer) individually to each tube ("hot start"). Round 1 PCR parameters were 35 cycles of 95°C for 1 min, 34°C for 30 sec, 72°C for 1 min, with a final 7 min extension at 72°C. 0.5  $\mu$ l of Round 1 PCR product was transferred to the Round 2 DNase-treated PCR mix (composition as for Round 1 but containing primers PAN-UI and PAN-DI) using the "hot start" procedure. Round 2 PCR parameters were as for Round 1 but using 30 cycles only and annealing at 45°C for 1 min.

- Cloning of PCR products

PCR products were cloned using the TA-cloning® kit (British Biotechnology) according to the manufacturer's recommendations.

- Sequencing

Sequencing reactions were performed using the "Prism ready reaction kit dye deoxyterminator cycle sequencing kit" (Applied Biosystems). Automatic sequence analysis was performed on an automatic sequencer (Applied Biosystems, 373 A).

- RT-PCR with ST1 primer sets

The first PCR round was performed directly from the cDNA reaction mixture according to the one-step RT-PCR technique described by Mallet et al. (28). This one-step RT-PCR procedure reduced the probability of airborne contamination when opening the tubes and transferring PCR reagents after an independent cDNA synthesis. RNA was extracted as previously from 2ml of plasma (snap-frozen in liquid nitrogen and stored at -80°C) or from a 500  $\mu$ l sucrose fraction with a total RT-activity above 10<sup>6</sup> dpm, and resuspended in 50  $\mu$ l of RNase-free water. For each RT-PCR reaction 10 $\mu$ l of RNA solution was incubated in a Perkin-Elmer 480 thermocycler, 15 min at 20°C with 1U of RNase-free DNASE 1 and 1.2  $\mu$ l of 10X DNASE buffer (50mM

Tris, 10mM MgCl<sub>2</sub> and 0,1mM DTT) containing 1U/μl of RNase-inhibitor (PROMEGA), and heated at 70°C for 10 min for DNase inactivation. The solution was placed on ice and mixed (in conditions preventing airborne dust/DNA contamination) with 88 μl of PCR mix containing: 1X taq buffer, 25 nM/tube dNTPs, 40pM/tube of each first round primer (ST1.1 upstream primer: 5' AGGAGTAAGGAAACCCAACGGAC 3' (SEQ ID NO:99); ST1.1 downstream primer: 5'TAAGAGTTGCACAAGTGCG 3' (SEQ ID NO:100)), 2.5U/tube of taq (Appligene) and 10U/tube of AMV-RT (Boehringer). Each tube was further incubated in a Perkin-Elmer 480 thermocycler for 10 min at 65°C, followed by 2h at 42°C for cDNA synthesis and 5 min at 95°C for inactivation of AMV-RT and DNA denaturation. First round parameters were 40 cycles of 95°C for 1 min, 53°C for 2.5 min, 72°C for 1 min, with a final extension of 10 min at 72°C. 10μl of the first round were transferred to the second round PCR mix previously treated at 20°C for 15 min with RNase-free DNase 1 (0.02U/μl) followed by DNase inactivation at 70°C for 10 min. This mix contained 1X taq buffer, 25 nM/tube dNTPs, 40pM/tube of each second round primers [ST1.2 upstream primer: 5'TCAGGGATAGCCCCCATCTAT3' (SEQ ID NO:101); ST1.2 downstream primer: 5'AACCCTTTGCCACTACATCAATTT3' (SEQ ID NO:102)] and 2.5U/tube of taq (Appligene). Second round parameters were 30 cycles of 95°C for 1 min, 53°C for 1.5 min, 72°C for 1 min, with a final extension of 8 min at 72°C. 20μl of this nested RT-PCR product were deposited on a 0,7% agarose gel containing ethidium bromide and exposed to UV light for the visualization of amplified products.

- Hybridisation analysis of PCR products: MSRV-pol detection by ELOSA

The protocol was essentially as previously described (21) but with the following modifications: Nunc Maxisorb microtitre plates were coated with 100 ng per well capture probe CpV1b (see Figure 44) either by passive

adsorption (21) or alternatively by using streptavidin coated plates and biotinylated CpV1b. Peroxidase-labelled detector probe DpV1 (see Figure 44) was used and the assay cut-off was defined as the mean of 4 negative controls plus 0.2 OD<sub>492</sub> units.

- RNA extraction, cDNA synthesis and PCR amplification from MS plasma samples :

Total RNA was extracted from human MS plasma by a guanidium method as described elsewhere (29). Total RNA extracted from 100 ul of plasma, were treated with RNase-free DNase I (0.1U/ $\mu$ l; Boehringer Mannheim, France) and reverse transcribed under the conditions recommended by the manufacturer, using Superscript reverse transcriptase (Gibco-BRL, FRANCE). The resulting cDNAs were amplified by semi-nested PCR through 35 cycles (94°C 1 min, 55°C 1 mn, 72°C 1 min 30 sec) and 72°C 8 min for a final extension. Three different fragments in the RT region were amplified by the following specific primers :

- in the protease (PRT) region, for the 1st and 2nd round of PCR, respectively, sense primer [5' TCC AGC AGC AGG ACT GAG GGT 3' (SEQ ID NO:103)] and antisense primers [5' CTG TCC GTT GGG TTT CCT TAC TCC T 3' (SEQ ID NO:104) / 5' GAC AGC AAA TGG GTA TTC CTT TCC 3' (SEQ ID NO:105)]

- in the fragment A of the RT region (Cf. Fig 46), for the 1st and 2nd round of PCR, respectively, sense primer [5' AGG AGT AAG GAA ACC CAA CGG ACA G 3' (SEQ ID NO:106)] and antisense primers [5' TGT ATA TAA TGG TCT GGC TAT TGG G 3' (SEQ ID NO:107) / 5' TTC GGC AGA AAC CTG TTA TGC CAA GG 3' (SEQ ID NO:108)]

- in the fragment B of the RT region (Cf. Fig. 46), for the 1st and 2nd round of PCR, respectively, sense primers [5' GGC TCT GCT CAC AGG AGA TTA GAT AC 3' (SEQ ID NO:109) / 5' AAA GGC ACC AGG GCC CTC AGT GAG GA 3' (SEQ ID NO:110)] and antisense primer 3'[5' GGT TTA AGA GTT GCA CAA GTG CGC AGT C 3' (SEQ ID NO:101)].

The amplified fragments were analysed on ethidium bromide-stained agarose gels, cloned in TA cloning vector (Invitrogen) and sequenced.

## 2) RESULTS

- 5 - Specific retroviral RNA is found in extracellular virions from MS patient-derived cell cultures and in MS patients' CSF.

Choroid plexus cells (4) (obtained post-mortem) and EBV-immortalized peripheral blood B-lymphocytes (30, 31) from MS patients gave rise to cultures expressing 100-120 nm viral particles associated with RT-activity similar to that of the original LM7 isolate (3). Similar cell-types from non-MS donors produced neither this RT-activity nor virions. All the 'infected' cultures were poorly and/or transiently productive and/or had a limited lifespan. Therefore, in order to analyse the genomic RNA present in the very limited quantity of extracellular virions, we used an RT-PCR approach to amplify, with degenerate primers, a conserved region of the pol gene present in all known retroviruses (12); the techniques based on this approach will be called "Pan-retro" RT-PCR. Extensive DNase treatment of samples and reagents was essential, because human DNA contains many endogenous retroviral elements amplifiable by this technique.

"Pan-retro" RT-PCR experiments were performed on sucrose-density gradient purified virions from supernatants of different types of cell cultures and their non-infected controls: (i) choroid plexus cells sampled post-mortem from MS brain (PLI-1), (ii) choroid plexus cells from non-MS brain autopsy, infected by co-culture with irradiated LM7 cells (LM7P), and (iii) identical non-infected choroid-plexus cells. "Early" B-cell lines obtained by spontaneous in vitro transformation of two EBV-seropositive individuals, (iv) one MS patient and (v) one non-MS control, were also analysed. Figure 43 illustrates the RT-activity in sucrose-gradient fractions obtained

from the B-cell cultures. The technique described by Shih et al. (12) was modified in a semi-nested RT-PCR protocol (27) using degenerate primers (Fig.2) and extensive DNase treatment. PCR amplifications were performed in London  
5 (Dpt of Virology, U.C.L.M.S.) on coded aliquots of the density gradient fractions. Blind and systematic cloning and sequencing of the PCR products were undertaken in an independent laboratory (bioMérieux, Lyon). After complete sequencing of 20 to 30 clones per sucrose gradient  
10 fraction, the codes were broken and results analysed in parallel with the RT-activity data.

Table 5 presents the distribution of sequences obtained from sucrose gradient fractions containing the peak of viral RT-activity in MS-derived cultures and also the  
15 sequences amplified from the corresponding RT-activity negative fractions of uninfected cultures. The predominant sequence detected in bands of the expected size (~140 bp) amplified in all the RT-activity positive fractions (but not in the RT-activity negative fractions) was different  
20 from known retroviruses and was designated MSRV-cpol. MSRV-cpol sequences exhibited partial homology (70-75%) with ERV9, a previously described endogenous retroviral sequence (18). A few ERV9 sequences (>90% homology with ERV9) were also present but clearly represented a minority  
25 of clones. In addition to typical pol sequences, numerous PCR artefacts (primer multimers, concatemers or single-primer amplifications) related to the use of degenerate primers and low-temperature annealing, were found in all samples (Table 5).

30 Figure 44 shows an alignment of a consensus sequence of MSRV-cpol with the corresponding VLPQG / YMDD region of diverse retroviruses. Figure 45 displays a phylogenetic tree based on the evolutionarily conserved amino acid sequences of both exogenous and endogenous retroviruses in this  
35 region. From this tree it can be seen that the pol gene of

MSRV is phylogenically related to the C-type group of oncovirinae.

A small scale study was performed to determine the prevalence of MSRV c-pol sequences in the CSF of patients with MS. Identification of MSRV-cpol in PCR products by cloning and sequencing is both laborious and time consuming. We therefore devised an enzyme-linked oligosorbent assay (ELOSA), using a capture probe (CpV1B) and a peroxidase-labelled detector probe (DpV1), for the rapid identification of MSRV-cpol sequences in 'Pan-retrovirus' PCR products (Figure 44). The specificity of this sandwich hybridisation-based assay for HMSRV-cpol was tested with both distantly related (HIV and MoMLV) and closely related (ERV9) pol sequences. No significant cross reactivity with such targets was observed despite the ability of the ELOSA to detect as little as 0.01 ng of MSRV-cpol DNA.

Cerebrospinal fluid (CSF) samples were available from 10 patients with MS and from 10 patients with other neurological disorders. Total RNA was extracted from CSF pellets, reverse transcribed and amplified as above. ELOSA analysis (Table 6) of the PCR products revealed MSRV-cpol sequences in 5 of the 10 MS patient samples but in none of the 10 samples from patients with other neurological diseases ( $P < 0.05$ ). The presence of MSRV-cpol did not appear to be correlated with age, sex or type of MS, but was seen in untreated patients only (5/6). No patient with immunosuppressive therapy was found positive (0/4). No correlation between MSRV-cpol detection and CSF cell count was observed.

- Cloning and sequencing a larger region of the pol gene

An independent identification of the MSRV genomic sequence was obtained by a non-PCR approach using RNA extracted from concentrated virions derived from 2,5 liters of LM7-infected sub-cultures of choroid plexus cells. A limited number of clones was obtained by direct

cloning of the cDNA, one of which (PSJ17) showed partial homology with ERV9 pol. Specific primers based on the MSRV-*cpol* region and on the PSJ17 clone, amplified a 740 bp fragment linking the two independent sequences in RNA  
5 extracted from purified virions. PSJ17 was localised on the 3' side of MSRV-*cpol*. Further sequence extension on the 5' side of MSRV-*cpol* and on the 3' side of PSJ17, was obtained using RT-PCR approaches on RNA from purified LM7-like virions produced in MS choroid plexus cultures (4).

10 In Figure 46, the nucleotide sequence corresponding to overlapping clones obtained by sequence extension in the pol gene is represented with the aminoacid translation corresponding to the putative open reading frames (ORFs) of the protease and of the reverse-  
15 transcriptase. The active site motifs of the protease (PRT) and of the reverse-transcriptase (RT) are underlined. In the C-terminal region of the RT sequence, the dispersed amino acid residues regularly present in retroviral RNase H domains, are also underlined.

20 - Non-degenerate primers detect MSRV-specific RNA in virions associated with the peak of RT-activity . and in in MS patients' plasma

PCR primers (ST1.1 primer set; positions 603-625/1732-1714, on Fig.4) based on overlapping clones in the pol  
25 gene, amplified a 1.15 kb segment of the RT region from several different isolates obtained from different MS patients. Nested primers (ST1.2; positions 869-889/1513-1490, on Fig.46) generated a 700 bp fragment (Figure 47) which was more easily visualised by ethidium bromide  
30 staining than the first round product generated by ST1.1. The specificity of PCR products was confirmed by stringent hybridisation with a peroxidase-labeled MSRV-*cpol* probe (Fig.44), using the ELOSA technique (21).

The ST1.1 and 2 primer set was used to detect  
35 extracellular MSRV RNA in human plasma, although non-optimal for this application. Figure 47 illustrates the

results of PCR amplification of cDNA derived from 2 MS patient and 2 control plasma samples tested in parallel with cDNA from the sucrose density gradient fractions of an MS choroid plexus isolate. Taq-sequencing of the 700 bp bands confirmed the presence of MSRV sequence. A very faint 700 bp band is also visible in fraction 10 which corresponds to the bottom of the tube where aggregated particles usually sediment. Control RT-PCR for cellular aldolase transcripts on plasma-derived RNA was negative, indicating that the results were not due to cellular RNA released by cell lysis during plasma separation. It should be noted that this PCR technique was not designed for epidemiological studies since its sensitivity is impaired by the length of the cDNA required (1.15 kb).

Non degenerate primers amplifying three fragments of the pol gene (the whole protease region, regions A and B of the reverse transcriptase; Cf. Fig. 46) were also used to confirm the presence of MSRV sequences in DNase-treated RNA from MS plasma. These fragments were amplified from the plasma of a further 4 MS patients with active disease. Sequence analysis confirmed that the PRT and RT regions were homologous (>95% and >90% respectively) to MSRV sequences previously obtained on culture virion. No such sequence were detected in plasma from healthy controls (n=4), tested in parallel with MS plasma.

### 3) DISCUSSION

#### - Phylogeny of MSRV

From the results of this study, it can be concluded that the virus previously referred to as "LM7" (3, 5, 26) possesses an RNA genome containing the MSRV pol sequences described here.

The conserved RT motif of both MSRV and ERV9 is two amino acids shorter than that of other retroviruses, apart from human foamy viruses which nonetheless have a functional RT. The potential ORF encompassing the entire PRT-RT



region is consistent with the virion-associated RT-activity detected in sucrose density gradients with infected culture supernatants. Moreover, since we have recently succeeded in expressing a recombinant protein  
5 from the sequence of MSRV protease cloned from MS plasma, we can confirm the reality of the potential PRT ORF. Similar cloning and expression of other sequences containing potential ORFs for MSRV proteins, is being undertaken to confirm their ability to encode enzymes and  
10 structural proteins of MSRV virions.

The phylogenic tree in Figure 45, based on the most conserved amino acid sequence in retroviruses (VLPQG...YXDD), shows that the MSRV pol gene is related to the C-type oncoviruses. Apart from ERV9, the closest known  
15 retroviral element is RTLV-H, a human endogenous sequence known to have a subtype with a functional pol gene (32). In the pol region, this phylogenic affiliation to C-type oncoviruses apparently contradicts our previous assumptions based on the general morphology of the  
20 particles observed by electron microscopy (EM), which were compatible with a B or D-type oncovirus (3, 5, 26). However, preliminary data on env sequences detected in MSRV virions, would suggest a greater phylogenic proximity to D-type. Such difference in phylogenies of the pol and  
25 env genes have been described in MPMV and suggest a recombinatorial origin in D-type retroviruses (33). D to C type morphological conversion is also possible since it has been reported that a single amino acid substitution in the gag protein can convert retrovirus morphology to that  
30 of a different type (34).

- Is MSRV an exogenous retrovirus sharing extensive homology with a related endogenous retrovirus family or an endogenous retrovirus producing extracellular virions?

Southern blot analysis with an MSRV pol probe  
35 under stringent conditions, showed hybridisation with a multicopy endogenous family (data not presented),

indicating the existence of endogenous elements more closely related to MSRV than ERV9 itself. Consequently, we were unable to look for a virion-specific provirus in MSRV-producing cells. In agreement with southern blot  
5 findings, PCR studies on genomic DNA showed multiple band amplification of MSRV-related endogenous sequences. Since pol is the most conserved retroviral gene, the sequence described here is the least suitable region to discriminate between exogenous and endogenous sequences.  
10 It is hoped that sequence information from other parts of the genome may permit such a discrimination, would it be on a tiny portion as has recently been demonstrated for the Jaagsiekte retrovirus (JSRV) of sheep (35). With such sequence data, it would then become possible to identify  
15 the MSRV-specific provirus in the genome of virion-producing cell cultures.

MSRV could represent a virion-producing exogenous member of an ERV9-like endogenous family, just as exogenous strains exist in the well-studied mouse mammary tumour  
20 virus (MMTV) and murine leukaemia virus (MuLV) retroviral families of mice, and also, in the JSRV retroviral family of sheep (36). Alternatively, it is also conceivable that the extracellular MSRV virions may be produced by a replication-competent endogenous provirus. Whether MSRV is  
25 exogenous or endogenous, conceptual similarities exist with the category of retroviruses represented by MuLV, MMTV and JSRV. Unlike defective endogenous elements, this category of agents are known to produce infectious and pathogenic virions, to cause neurological disease (37),  
30 solid tumours / leukaemias (36, 38) and to express "endogenous superantigens" (39, 40). Furthermore, in MuLV infections, the genetic endogenous retroviral background of the mouse strain can determine susceptibility or resistance to disease (39, 41). Indeed, such interactions  
35 between an infectious retrovirus and its endogenous counterpart may be relevant in the pathogenesis of MS,

since endogenous retroviral genotypes are not identical in all individuals. A genetic control due to related endogenous retroviral genotypes could therefore contribute to the known hereditary susceptibility to MS (43), if MSRV  
5 does indeed play an active role in this disease.

Elsewhere, the data in Table 5 suggest that ERV9 elements may be co-expressed, possibly via trans-activation in infected cells, and give rise to heterologous RNA packaging in MSRV virions. Such heterologous packaging is  
10 known to occur in other retroviral systems (42).

- A role for the numerous common viruses previously evoked in MS ?

Among the numerous reports of viruses putatively involved in the aetiopathogenesis of MS, a significant  
15 proportion focus on two viral families, the paramyxoviridae and the herpesviridae. Regarding the paramyxoviridae, the key observation is of a frequently increased antibody titer to measles virus in MS patients essentially directed, in CSF, against measles fusion  
20 protein (44). The existence of aminoacid similarities between conserved domains of the fusion proteins of paramyxoviridae and the transmembrane protein of retroviruses (45), may explain this observation if antigenic cross-reactivity between these two proteins  
25 occurred.

With regard to the herpesvirus family, the involvement of Epstein-Barr Virus (EBV), Herpes Simplex Virus type 1 (HSV-1) and, most recently, Human Herpes Virus 6 (HHV-6) has been proposed (31, 46, 47). From our previous studies  
30 and from those of other groups, it appears that herpesviruses may play an important role in MSRV expression: we have shown that HSV-1 immediate-early ICP0 and ICP4 proteins can transactivate MSRV/LM7 in vitro (6) and Haahr et al. have proposed an important  
35 epidemiological role for EBV, as a co-factor in MS, triggering retrovirus reactivation (31). The recent

description by Challoner et al. (47) showing significant expression of HHV6 proteins in MS plaques may also suggest a similar role for HHV6 in the brain.

#### 5 EXAMPLE 19 : MSRV GENOME DETECTION TECHNIQUE

Following 0.4  $\mu$ m filtration to remove cellular debris and RNase digestion to remove residual non-encapsidated RNA, serum was processed to extract viral RNA by means of adsorption to a silica matrix. Viral RNA was  
10 subjected to DNase digestion, then a combined reverse transcription-PCR (RT-PCR) reaction was performed using primers PTPol-A (sense: 5'xxxx3', SEQ ID NO:183) and PTPol-F (antisense: 5'xxxx3', SEQ ID NO:184). A second round of amplification with nested primers PTPol-B (sense:  
15 5'xxxx3', SEQ ID NO:185) and PTPol-E (antisense: 5'xxxx3', SEQ ID NO:186) generated a 435 bp PCR product which was identified by gel electrophoresis. The specificity of each product was confirmed by dideoxy sequencing. Control reactions without reverse transcriptase were performed to  
20 ensure that the products were derived from viral RNA. In addition, to exclude the possibility that the extracted viral RNA might be contaminated with host cell derived nucleic acids, aliquots were tested by nested PCR for the presence of pyruvate dehydrogenase (PDH) DNA and RNA.  
25 Samples which generated a signal in either the PDH or the "no-RT" PCR assays were excluded from the analysis.

Sera from patients with clinically active MS and controls were amplified by RT-PCR and sequenced. Virion associated MSRV-RNA was detected in the serum of 10 of 19  
30 (53%) patients with MS but in only 3 of 44 controls without MS (P=0.0001). The control group consisted of 8 patients (all MSRV-RNA negative) with rheumatological disorders and 36 healthy adults. MSRV-RNA titres in both MS patients and controls were apparently low because even  
35 moderate dilution of sera (<10 fold) caused loss of signal.

In MS patients, detection of MSRV-RNA was not associated with age, sex, disease duration, or MS type, however a significant negative correlation with treatment was observed. 26 serum samples were obtained from the 19 patients ; 100% of the sera from untreated patients contained detectable MSRV-RNA whereas it was detectable in only 4 of 19 samples (21%) obtained during treatment with corticosteroids and/or azathioprine ( $P=0.001$ ).

The reason for the apparent loss of virion associated MSRV-RNA during immunosuppressive treatment is unknown but the finding is in agreement with the previous observations on the detection of MSRV in cerebrospinal fluid.

15

TABLE 7

DETECTION OF VIRION ASSOCIATED MSRV-RNA IN MS UNTREATED PATIENTS & CONTROLS

	Positive	Negative	Total	% Positive
Controls without MS <sup>a</sup>	3 <sup>b</sup>	41	44	7%
MS sera untreated at time of sampling	7	0	7	100%

<sup>a</sup> The control group consisted of 8 patients with miscellaneous non-MS disorders and 36 healthy adults.

<sup>b</sup> The detection of MSRV RNA in plasma of a few controls in conditions which select virion-packaged RNA, is consistent with the knowledge that a virus associated with MS should be present in a minor proportion of apparently healthy population. Indeed, such individuals can be either healthy carriers or be in the pre-clinical (or sub-clinical) phase of the disease which can last for years.

30

## METHOD :

- Modified SNAP RNA extraction with filtration and RNase digestion

(All centrifugations are at room temperature)

5 Up to 500 microlitres of serum is filtered using 0.45 micron spin filters (Nanosep MF from Flowgen Catalogue No. U3-0126 Ref. ODM45). The serum is spun for 5 min at 130,000 g (or for further 10 min if necessary).

10 150 microlitres of filtered serum is incubated with 10 units RNase One (Promega Catalogue No.M4261) for 30 min at 37°C.

The 150 microlitres was then extracted using the SNAP RNA extraction kit (Invitrogen) as below:

- 10 micrograms of poly A RNA was added to the  
15 450 microlitres of Binding Buffer to act as a carrier ; this was then added to the serum and mixed by inversion 6 times ; 300 microlitres of propan-2-ol was then added and mixed by inversion 10 times ; 500 microlitres was transferred to the SNAP column and spun at 1300 g for  
20 1 min and the flow-through discarded ; the remainder was then added to the SNAP column and spun at 1300 g for 1 min and the flow-through discarded ; the column was then washed with 600 microlitres of Super wash and the flow-through discarded ; the column was then washed with 600  
25 microlitres of 1x RNA wash and the flow-through discarded ; this wash was repeated with a 2 min 1300 g spin and the flow-through discarded ; the bound nucleic acid was then eluted by incubating with 135 microlitres of RNase free water for 5 min and spun at 1300 g for 1 min.

30 - 15 microlitres of 10x DNase buffer and 3 microlitres (30 units) of DNase I, RNase free (Boehringer Mannheim Cat. No. 776 785) was added and incubated for 30 min at 37°C ; 450 microlitres of Binding Buffer was added and mixed by inversion 6 times ; 300 microlitres of  
35 propan-2-ol was then added and mixed by inversion 10 times ; 500 microlitres was transferred to the SNAP column

and spun at 1300 g for 1 min and the flow-through discarded ; the remainder was then added to the SNAP column and spun at 1300 g for 1 min and the flow-through discarded ; the column was then washed with 600  
5 microlitres 1x RNA wash and the flow-through discarded ; this wash was repeated with a 2 min 1300 g spin and the flow-through discarded ; the bound nucleic acid was then eluted by incubating with 105 microlitres of RNase free water for 5 min and spun at 1300 g for 1 min.

10

- Titan RT-PCR

RT-PCR was performed using the Titan one tube RT-PCR system (Boehringer Mannheim Cat. No. 1 855 476) 25 microlitres of RNA was used in the combined RT-PCR  
15 reaction. The total reaction volume was 50 microlitres. Promega rRNasin (10 units) was the RNase inhibitor used. 170 ng of primers SEQ ID NO:183 and SEQ ID NO:184, respectively, were used. A single master mix was prepared and the sample RNA added last. This was performed at room  
20 temperature, not on ice.

The RT step consisted of two sequential 30 min incubations at 50°C and then 60°C. This was immediately followed by the PCR which had the following steps.

- \* Initial denaturation of template at 94°C for 2 min,
- 25 \* 40 cycles of 94°C for 30 seconds ; 60°C for 30 seconds ; 68°C for 45 seconds,
- \* 1 cycle of 68°C for 7 min.

The second round PCR was performed using the Expand long template PCR system (Boehringer Mannheim Cat. No. 1681 842). 0.5 microlitres of the RT-PCR mix was added  
30 to 25 microlitres of the round 2 PCR mix. Buffer No. 3 and 50 ng of primers B and E were used. The PCR had the following steps:

- \* 5 cycles of 94°C for 30 seconds, 60°C for 30 seconds.,
- 35 68°C for 45 seconds,
- \* 1 cycle of 68°C for 7 min.

The PCR products were then run on a 2% agarose gel.

The no RT controls were performed using "Expand" PCR system for both rounds. The first round was 40 cycles and the second round 20 cycles.

As a positive control a DNA dilution series was used in both the RT-PCR and the "no RT" PCR. For a result to be valid the RT-PCR and "no-RT" PCRs had to have detected DNA equivalent to between 1 and 0.1 cells.

The analysis of PCR products of an approximately 435 bp fragment in the pol region is shown in Table 8.

TABLE 8  
ANALYSIS OF PCR PRODUCTS WITH ORF \*

Exp	Disease	Clone	ORF	Fragment (bp)	AA-RT Motif Site
46-7	MS	1	+	429	YGDD
		5	+	429	YGDD
		8	+	429	YGDD
68-1	MS	41	+	438	YMDD
		42	+	438	YMDD
		43	+	438	YMDD

\* Defective RNA can also be present in circulating virions, since the fidelity of the MSRV reverse transcriptase appears to be low and since recombination events with related endogenous elements can occur. It is then obvious that the intra- and inter- patients variability can be greater than that illustrated in this example, because of these encapsidated defective MSRV RNA copies.

Table 9 which data have been determined from the alignments of Figures 49 to 53, shows a variability :



- between the clones obtained from the same patient plasma sample in the same PCR amplification experiment ; this means that the patient possesses a virion population which comprises different MSRV variants at a given time,
- 5 - between the sequenced variant populations from different patients ; this means that the variants differ from a patient to another patient.

TABLE 9

10 Degree of identity (percentage) between nucleotide sequences and between peptide sequences, by direct comparison of said sequences (see Figures 49-53)

Patient	68-1	46-7
Nucleotide sequences	between SEQ ID NO:169 and MSRV-pol (SEQ ID NO:1) 90,4 % <sup>b</sup>  92,3 % <sup>a</sup>  SEQ ID NOS:170, 171, 172 between them 98,6 % <sup>b</sup>  98,7 % <sup>a</sup>	between SEQ ID NO:176 and MSRV-pol (SEQ ID NO:1) 82,5 % <sup>a</sup>  84 % <sup>b</sup>  SEQ ID NOS:177, 178, 179 between them 94,5 % <sup>a</sup>  95,1 % <sup>b</sup>
Peptide sequences	between SEQ ID NOS:173, 174, 175 and SEQ ID NO: 81 %  SEQ ID NOS:173, 174, 175 between them 97 %	between SEQ ID NOS:180, 181, 182 and SEQ ID NO: 73,5 %  SEQ ID NOS:180, 181, 182 between them 89 %

- 15 a) this percentage is determined on the basis of sequences excluding the primers  
 b) this percentage is determined on the basis of sequences including the primers.
- 20 From Figures 53A and 53B, the variability between tested patients sequences can be determined :

116

- between SEQ ID NO:169 and SEQ ID NO:176 : 16,5 %<sup>a</sup> and 14,8 %<sup>b</sup>
- between the peptide sequences obtained from SEQ ID NO:169 and SEQ ID NO:176 : 20 %.

5

Four microorganisms are mentioned in the specification page 3 lines 15-26 and they are identified below. They have all been deposited with the ECACC\*, in accordance with the provisions of the Budapest Treaty.

10

- LM7PC deposited on 22nd July 1992 under No. 92072201,
  - PLI-2 deposited on 8th January 1993 under No. 93010817,
  - POL-2 deposited on 22nd July 1992 under No. V92072202,
- and

- 15 - MS7PG deposited on 8th January 1993 under No. V93010816.

\* ECACC : European Collection of Animal Cell Cultures  
Vaccine Research and Production Laboratory  
Public Health Laboratory Service  
Centre of Applied Microbiology and Research  
Porton Down  
Salisbury, Wiltshire SP4 0JG  
United Kingdom

20

25

## BIBLIOGRAPHY

- (1) Norrby E., Prog. Med. Virol., 1978; 24, 1-39.
- (2) Johnson R.T., "Handbook of clinical neurology, 47 Demyelinating diseases", Vinken P. and Bruyn G.W., eds. 5 Amsterdam, Elsevier Science Publishing, 1985, 319-336.
- (3) Perron H. et al., Res. Virol. 1989, 140, 551-561.
- (4) Perron H. et al., "Current concepts in multiple sclerosis" Wiethölter et al., eds. Amsterdam, Elsevier, 1991, 111-116.
- 10 (5) Perron H. et al., The Lancet 1991, 337, 862-863.
- (6) Perron H. et al., J. Gen. Virol. 1993, 74, 65-72.
- (7) Fields and Knipe, Fundamental Virology 1986, Rev Press N.Y.
- (8) Nielsen P.E. et al., Science 1991; 254, 1497-1500.
- 15 (9) Maniatis et al., Molecular Cloning, Cold Spring Harbour, 1982.
- (10) Southern. E.M., J. Mol. Biol. 1975, 98, 503.
- (11) Dunn A.R. and Hassel J.A., Cell 1977, 12, 23,
- (12) Shih et al., J. Virol. 1989, 63, 64-75.
- 20 (13) Perron H. et al., Res. Vir. 1992, 143, 337-350.
- (14) Meyerhans et al., Cell 1989, 58, 901-910.
- (15) Linial M.L. and Miller A.D., "Current topics in microbiology and immunobiology. Retroviruses, strategies of replication" vol. 157, 125-152; Swanstrom R. and Vogt 25 P.K., editors, Springer-Verlag, Heidelberg 1990.
- (16) Lori F. et al., J. Virol. 1992, 66, 5067-5074.
- (17) Sambrook J., Fritsch E.F. and Maniatis T., Molecular cloning, a laboratory manual. Cold Spring Harbour Laboratory Press, 1989.
- 30 (18) La Mantia et al., Nucleic Acids Research 1991, 19, 1513-1520.
- (19) Gonzales-Quintial R, Baccala R, Pope R M and Theofilopoulos N, J. Clin. Invest, Vol. 97, Number 5, pp1335-1343, 1996.
- 35 (20) Chomzynski P. and N. Sacchi, Analytical Biochemistry 1987, 162, 156-159.

- (21) F. Mallet et al., Journal of Clinical Microbiology 1993; 31, 1444-1449.
- (22) G. Barany and R.B. Merrifield, 1980, In the Peptides, 2, 1-284, Gross E and Meienhofer J, Eds.,  
5 Academic Press, New York.
- (23) Poser et al., Gbers G.C. eds. The diagnosis of multiple sclerosis Thieme Stratton Inc, New York 1984: 225-229.
- (24) La Mantia et al., Nucleic Acid Research 1989, 17,  
10 5913-22.
- (25) PLAZA, A; KONO, D.H.; THEOFILOPOULOS, A.N. NEW HUMAN  $V\beta$  GENES AND POLYMORPHIC VARIANTS. J. Imm; 147(12): 4360-4365, 1991.
- (26) H. Perron, In vitro transmission and antigenicity of  
15 a retrovirus isolated from multiple sclerosis, Res. Virol. 143, 337-350 (1992).
- (27) J. Garson et al., Development of a "Pan-retrovirus" detection system for multiple sclerosis studies. Acta Neurol. Scand. (in Press).
- 20 (28) F. Mallet, G. Oriol, C. Mary, B. Verrier and B. Mandrand. Continuous RT-PCR and taq DNA polymerase : Characterization and comparison to uncoupled procedures. Biotechniques 18, 678-687 (1995).
- (29) R. Baccala, D.H. Kono, S. Walker, R.S. Balderas and  
25 Theophilopoulos. Genomically imposed and somatically modified human thymocyte  $v\beta$  gene repertoires. Proc. Natl. Acad. Sci. USA (1991) 88, 2908.
- (30) Haahr S., Koch-Henriksen N., Møller-Larsen, A. Eriksen L.S. & Andersen H.M.K. Increased risk of multiple  
30 sclerosis after late Epstein-Barr virus infection: a historical prospective study. Multiple Sclerosis 1, 73-77 (1995).
- (31) Haahr S et al. A putative new retrovirus associated with multiple sclerosis and the possible involvement of  
35 Epstein-Barr virus in this disease. Ann. NY Acad. Science. 724, 148-156 (1994).

- (32) Wilkinson D.A., Goodchild N.L., Saxton T.M., Wood S. & Mager D.L. Evidence for a functional subclass of the RTLV-H family of human endogenous retrovirus-like sequences. *J. Virol.* 67, 2981-2989 (1993).
- 5 (33) Sonigo P., Barker C., Hunter E. and Wain-Hobson S. Nucleotide sequence of Mason-Pfizer monkey virus: an immunosuppressive D-type retrovirus. *Cell* 45, 375-85 (1986).
- (34) Rhee S.S., and Hunter E. A single amino acid  
10 substitution within the matrix protein of a D-type retrovirus converts its morphogenesis to that of a C-type retrovirus. *Cell* 63, 77-86 (1990).
- (35) Bai J., Zhu R.Y., Stedman K., Cousens C., Carlson J., Sharp J.M. and DeMartini J.C. Unique long terminal repeat  
15 U3 sequences distinguish exogenous Jaagsiekte sheep retroviruses associated with ovine pulmonary carcinoma from endogenous loci in the sheep genome. *J. Virol.* 70, 3159-3168 (1996).
- (36) Palmarini M., Cousens C., Dalziel R.G., Bai J.,  
20 Stedman K., DeMartini J.C. and Sharp J.M. The exogenous form of Jaagsiekte retrovirus is specifically associated with a contagious lung cancer of sheep. *J. Virol.* 70, 1618-1623 (1996).
- (37) Portis J.L. Wild mouse retrovirus : pathogenesis. in  
25 "Retrovirus infections of the nervous system". Oldstone M.B.A. and Koprowsky H. Eds. Current topics in microbiology and immunology, n°160, p. 11-27. (Springer-Verlag, Berlin 1990).
- (38) Gardner M.B., Chivi A., Dougherty M.F., Casagrande J  
30 & Estes J.D. Congenital transmission of murine leukaemia virus from wild mice prone to development of lymphoma and paralysis. *J. Natl. Cancer Inst.* 62, 63-69 (1979).
- (39) Marrack P., Kushnir E. & Kappler J. A maternally inherited superantigen encoded by a mammary tumor virus.  
35 *Nature* 349, 524-526 (1991).

- (40) Hügin A.W., Vacchio M.S. & Morse H.C. A virus-encoded superantigen in a retrovirus-induced immunodeficiency syndrome of mice. *Science* 252, 424-427 (1991).
- (41) Gardner M.B. Genetic resistance to a retroviral neurologic disease in wild mice, in "Retrovirus infections of the nervous system" Oldstone M.B.A. and Koprowsky H. Eds. *Current topics in microbiology and immunology*, n° 160, p. 3-10. (Springer-Verlag, Berlin 1990).
- (42) Linial M.L. & Miller A.D. Retroviral RNA packaging: sequence requirements and implications, in "Retroviruses-strategies of replication" Swanstrom R. & Vogt P.K. Eds. *Current topics in microbiology and immunology*, n° 157, p. 125-152. (Springer-Verlag, Berlin 1990).
- (43) Bell J.I. and Lathrop G.M. Multiple loci for multiple sclerosis. *Nature Genetics* 13, 377-78 (1996).
- (44) Dhib-Jalbut S., Lewis K., Bradburn E., McFarlin D.E. and McFarland H.F. Measles virus polypeptide-specific antibody profile in multiple sclerosis. *Neurology*, 1990; 40 : 430-435.
- (45) Gonzalez-Scarano F., Waxham M.N., Ross A.M. and Hoxie J.A. Sequence similarities between human immunodeficiency virus gp41 and Paramyxovirus fusion proteins. *AIDS Res. Hum. Retrov.* 1987 ; 3 : 245-252.
- (46) Bergström, T., Andersen, O. & Vahlne A. (1989). Isolation of herpes virus type 1 during first attack of multiple sclerosis. *Annales Neurology* 26 , 283-285.
- (47) Challoner P.B. et al. Plaque-associated expression of human herpesvirus 6 in multiple sclerosis. *Proc. Natl. Acad. Sci. USA* 92, 7440-7444 (1995).
- (48) A. Gessain et al; Antibodies to Human T-Lymphotropic Virus type-I in patients with tropical spastic paraparesis. *Lancet* 2, 407-410 (1985).
- (49) H. Perron, J.A. Garson, F. Bedin et al., Molecular identification of a novel retrovirus repeatedly isolated

121

from patients with multiple sclerosis. Proc. Nat. Acad.  
Sci. USA 94:7583-7588 (1997).

122

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

5           (i) APPLICANT:    BIO MERIEUX

          (ii) TITLE OF THE INVENTION: VIRAL MATERIAL AND NUCLEOTIDE  
FRAGMENTS ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC,  
10   PROPHYLACTIC AND THERAPEUTIC PURPOSES

          (iii) NUMBER OF SEQUENCES: 160

15           (iv) CORRESPONDENCE ADDRESS:

          (A) ADDRESSEE: CABINET GERMAIN & MAUREAU

          (B) STREET: 12 rue Boileau

          (C) CITY: LYON

          (D) COUNTRY: FRANCE

20           (E) ZIP: 69006

          (v) COMPUTER READABLE FORM:

          (A) MEDIUM TYPE: Floppy disk

          (B) COMPUTER: IBM PC compatible

25           (C) OPERATING SYSTEM: PC-DOS/MS-DOS

          (D) SOFTWARE: PatentIn Release #1.0, Version #1.30

          (vi) CURRENT APPLICATION DATA:

          (A) APPLICATION NUMBER:

30           (B) FILING DATE:

          (viii) ATTORNEY/AGENT INFORMATION:

          (A) NAME: Dominique GUERRE

          (B) REGISTRATION NUMBER:

35           (C) REFERENCE/DOCKET NUMBER: MD/B05B2679



123

## (ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 4 72 69 84 30

(B) TELEFAX: 4 72 69 84 31

## 5 (2) INFORMATION FOR SEQ ID NO: 1:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1158 base pairs

(B) TYPE: nucleotide

10 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## 15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
CCCTTTGCCA CTACATCAAT TTTAGGAGTA AGGAAACCCA ACGGACAGTG GAGGTTAGTG 60
CAAGAACTCA GGATTATCAA TGAGGCTGTT GTTCCTCTAT ACCCAGCTGT ACCTAACCCT 120
TATACAGTGC TTTCCCAAAT ACCAGAGGAA GCAGAGTGGT TTACAGTCCT GGACCTTAAG 180
20 GATGCCTTTT TCTGCATCCC TGTACGTCCT GACTCTCAAT TCTTGTTTGC CTTTGAAGAT 240
CCTTTGAACC CAACGTCTCA ACTCACCTGG ACTGTTTTAC CCCAAGGGTT CAGGGATAGC 300
CCCCATCTAT TTGGCCAGGC ATTAGCCCAA GACTTGAGTC AATTCTCATA CCTGGACACT 360
CTTGTCCTTC AGTACATGGA TGATTACTT TTAGTCGCCC GTTCAGAAAC CTTGTGCCAT 420
CAAGCCACCC AAGAACTCTT AACTTTCCTC ACTACCTGTG GCTACAAGGT TTCCAAACCA 480
25 AAGGCTCGGC TCTGCTCACA GGAGATTAGA TACTNAGGGC TAAAATTATC CAAAGGCACC 540
AGGGCCCTCA GTGAGGAACG TATCCAGCCT ATACTGGCTT ATCCTCATCC CAAAACCCTA 600
AAGCAACTAA GAGGGTTCCT TGGCATAACA GGTTTCTGCC GAAAACAGAT TCCCAGGTAC 660
ASCCCAATAG CCAGACCATT ATATACACTA ATTANGGAAA CTCAGAAAGC CAATACCTAT 720
TTAGTAAGAT GGACACCTAC AGAAGTGGCT TTCCAGGCCC TAAAGAAGGC CCTAACCCAA 780
30 GCCCCAGTGT TCAGCTTGCC AACAGGGCAA GATTTTCTT TATATGCCAC AGAAAAAACA 840
GGAATAGCTC TAGGAGTCCT TACGCAGGTC TCAGGGATGA GCTTGCAACC CGTGGTATAC 900
CTGAGTAAGG AAATTGATGT AGTGGCAAAG GGTGGCCTC ATNGTTTATG GGTAATGGNG 960
GCAGTAGCAG TCTNAGTATC TGAAGCAGTT AAAATAATAC AGGGAAGAGA TCTTNCTGTG 1020
TGGACATCTC ATGATGTGAA CGGCATACTC ACTGCTAAAG GAGACTTGTG GTTGTGAGAC 1080
35 AACCATTTAC TTAANTATCA GGCTCTATTA CTTGAAGAGC CAGTGCTGNG ACTGCGCACT 1140
TGTGCAACTC TTAAACCC 1158
```

124

## (2) INFORMATION FOR SEQ ID NO: 2:

## (i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 297 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## 10 (ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

CCCTTTGCCA CTACATCAAT TTTAGGAGTA AGGAAACCCA ACGGACAGTG GAGGTTAGTG 60  
15 CAAGAACTCA GGATTATCAA TGAGGCTGTT GTTCCTCTAT ACCCAGCTGT ACCTAACCCCT 120  
TATACAGTGC TTTCCCAAAT ACCAGAGGAA GCAGAGTGGT TTACAGTCCT GGACCTTAAG 180  
GATGCCTTTT TCTGCATCCC TGTACGTCCT GACTCTCAAT TCTTGTTTGC CTTTGAAGAT 240  
CCTTTGAACC CAACGTCTCA ACTCACCTGG ACTGTTTTAC CCCAAGGGTT CAAGGGA 297

20

## (2) INFORMATION FOR SEQ ID NO: 3:

## (i) SEQUENCE CHARACTERISTICS:

- 25 (A) LENGTH: 85 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## 30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GTTTAGGGAT ANCCCTCATC TCTTTGGTCA GGTACTGGCC CAAGATCTAG GCCACTTCTC 60  
AGGTCCAGSN ACTCTGTYCC TTCAG 85

35

125

## (2) INFORMATION FOR SEQ ID NO: 4:

## (i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 86 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GTTCAGGGAT AGCCCCCATC TATTTGGCCA GGCAGTAGCT CAATACTTGA GCCAGTTCTC 60  
ATACCTGGAC AYTCTYGTCC TTCGGT 86

15

## (2) INFORMATION FOR SEQ ID NO: 5:

## (i) SEQUENCE CHARACTERISTICS:

- 20 (A) LENGTH: 85 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## 25 (ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GTTCARRGAT AGCCCCCATC TATTTGGCCW RGYATTAGCC CAAGACTTGA GYCAATTCTC 60  
30 ATACCTGGAC ACTCTTGTCC TTYRG 85

## (2) INFORMATION FOR SEQ ID NO: 6:

## 35 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 85 base pairs

126

- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

GTTCAGGGAT AGCTCCCATC TATTTGGCCT GGCATTAACC CGAGACTTAA GCCAGTTCTY 60  
10 ATACGTGGAC ACTCTTGTCC TTTGG 85

(2) INFORMATION FOR SEQ ID NO: 7:

- 15 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 111 base pairs
  - (B) TYPE: nucleotide
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

25

GTGTTGCCAC AGGGGTTTAR RGATANCYCY CATCTMTTTG GYCWRGYAYT RRCYCRKAY 60  
YTRRGYCAVT TCTYAKRYSY RGSNAYTCTB KYCCTTYRGT ACATGGATGA C 111

30 (2) INFORMATION FOR SEQ ID NO: 8:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 645 base pairs
  - (B) TYPE: nucleotide
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

35

127

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

5  
TCAGGGATAG CCCCCATCTA TTTGGCCAGG CATTAGCCCA AGACTTGAGT CAATTCTCAT 60  
ACCTGGACAC TCTTGTCTT CAGTACATGG ATGATTTACT TTTAGTCGCC CGTTCAGAAA 120  
CCTTGTGCCA TCAAGCCACC CAAGAACTCT TAACTTTCCT CACTACCTGT GGCTACAAGG 180  
TTTCCAAACC AAAGGCTCGG CTCTGCTCAC AGGAGATTAG ATACTNAGGG CTAAAATTAT 240  
10 CCAAAGGCAC CAGGGCCCTC AGTGAGGAAC GTATCCAGCC TATACTGGCT TATCCTCATC 300  
CCAAAACCCT AAAGCAACTA AGAGGGTTCC TTGGCATAAC AGGTTTCTGC CGAAAACAGA 360  
TTCCCAGGTA CASCCTAATA GCCAGACCAT TATATACACT AATTANGGAA ACTCAGAAAG 420  
CCAATACCTA TTTAGTAAGA TGGACACCTA CAGAAGTGGC TTTCCAGGCC CTAAAGAAGG 480  
CCCTAACCCA AGCCCAGTG TTCAGCTTGC CAACAGGGCA AGATTTTCTT TTATATGCCA 540  
15 CAGAAAAAAC AGGAATAGCT CTAGGAGTCC TTACGCAGGT CTCAGGGATG AGCTTGCAAC 600  
CCGTGGTATA CCTGAGTAAG GAAATTGATG TAGTGGCAA GGGTT 645

(2) INFORMATION FOR SEQ ID NO: 9:

20

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 741 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

25

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

30

CAAGCCACCC AAGAACTCTT AAATTTCTC ACTACCTGTG GCTACAAGGT TTCCAAACCA 60  
AAGGCTCAGC TCTGCTCACA GGAGATTAGA TACTTAGGGT TAAAATTATC CAAAGGCACC 120  
AGGGGCCTCA GTGAGGAACG TATCCAGCCT ATACTGGGT ATCCTCATCC CAAAACCTA 180  
AAGCAACTAA GAGGGTTCCT TAGCATGATC AGGTTTCTGC CGAAAACAAG ATTCCCAGGT 240  
35 ACAACCAAAA TAGCCAGACC ATTATATACA CTAATTAAGG AAACCTCAGAA AGCCAATACC 300  
TATTTAGTAA GATGGACACC TAAACAGAAG GCTTTCCAGG CCCTAAAGAA GGCCCTAACC 360

128

CAAGCCCCAG TGTTTCAGCTT GCCAACAGGG CAAGATTTTT CTTTATATGG CACAGAAAAA 420  
ACAGGAATCG CTCTAGGAGT CCTTACACAG GTCCGAGGGA TGAGCTTGCA ACCCGTGGCA 480  
TACCTGAATA AGGAAATTGA TGTAAGTGGCA AAGGGTTGGC CTCATNGTTT ATGGGTAATG 540  
GNGGCAGTAG CAGTCTNAGT ATCTGAAGCA GTTAAAATAA TACAGGGAAG AGATCTTNCT 600  
5 GTGTGGACAT CTCATGATGT GAACGGCATA CTCACTGCTA AAGGAGACTT GTGGTTGTCA 660  
GACAACCATT TACTTAANTA TCAGGCTCTA TTAAGTGAAG AGCCAGTGCT GNGACTGCGC 720  
ACTTGTGCAA CTCTTAAACC C 741

## 10 (2) INFORMATION FOR SEQ ID NO: 10:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 93 base pairs  
(B) TYPE: nucleotide  
15 (C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

TGGAAAGTGT TGCCACAGGG CGCTGAAGCC TATCGCGTGC AGTTGCCGGA TGCCGCCTAT 60  
AGCCTCTACA TGGATGACAT CCTGCTGGCC TCC 93

25

## (2) INFORMATION FOR SEQ ID NO: 11:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 96 base pairs  
30 (B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

35

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

129

TTGGATCCAG TGYTGCCACA GGGCGCTGAA GCCTATCGCG TGCAGTTGCC GGATGCCGCC 60  
TATAGCCTCT ACGTGGATGA CCTSCTGAAG CTTGAG 96

5

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 748 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

TGCAAGCTTC ACCGCTTGCT GGATGTAGGC CTCAGTACCG GNGTGCCCCG CGCGCTGTAG 60  
TTCGATGTAG AAAGCGCCCG GAAACACGCG GGACCAATGC GTCGCCAGCT TGCGCGCCAG 120  
20 CGCCTCGTTG CCATTGGCCA GCGCCACGCC GATATCACCC GCCATGGCGC CGGAGAGCGC 180  
CAGCAGACCG GCGGCCAGCG GCGCATTCTC AACGCCGGGC TCGTCGAACC ATTCGGGGGC 240  
GATTTCCGCA CGACCGCGAT GCTGGTTGGA GAGCCAGGCC CTGGCCAGCA ACTGGCACAG 300  
GTTCAGGTAA CCCTGCTTGT CCCGCACCAA CAGCAGCAGG CGGGTCGGCT TGTCGCGCTC 360  
GTCGTGATTG GTGATCCACA CGTCAGCCCC GACGATGGGC TTCACGCCCT TGCCACGCGC 420  
25 TTCCTGTAG ANGCGACCA GCCCGAAGGC ATTGGCGAGA TCGGTCAGCG CCAAGGCGCC 480  
CATGCCATCT TTGGCGGCAG CCTTGACGGC ATCGTCGAGA CGGACATTGC CATCGACGAC 540  
GGAATATTCTG GAGTGGAGAC GGAGGTGGAC GAAGCGCGGC GAATTCATCC GCGTATTGTA 600  
ACGGGTGACA CCTTCCGCAA AGCATTCCGG ACGTGCCCGA TTGACCCGGA GCAACCCCGC 660  
ACGGCTGCGC GGGCAGTTAT AATTTCCGGT TACGAATCAA CGGGTTACCC CAGGGCGCTG 720  
30 AAGCCTATCG CGTGCAGTTG CCGGATGC 748

(2) INFORMATION FOR SEQ ID NO: 13:

35

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs

130

- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

GCATCCGGCA ACTGCACG

18

10

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

15

(ii) MOLECULE TYPE: cDNA

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

GTAGTTCGAT GTAGAAAGCG

20

25

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

30

(ii) MOLECULE TYPE: cDNA

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:



131

GCATCCGGCA ACTGCACG

18

## 5 (2) INFORMATION FOR SEQ ID NO: 16:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleotide
- 10 (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## 15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

AGGAGTAAGG AAACCCAACG GAC

23

## 20 (2) INFORMATION FOR SEQ ID NO: 17:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleotide
- 25 (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## 30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

TAAGAGTTGC ACAAGTGCG

19

## 35 (2) INFORMATION FOR SEQ ID NO: 18:

132

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

5 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

10

TCAGGGATAG CCCCCATCTA T

21

## (2) INFORMATION FOR SEQ ID NO: 19:

15

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

20 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

25

AACCCTTTGC CACTACATCA ATTT

24

## (2) INFORMATION FOR SEQ ID NO: 20:

30

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 15 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

35 (D) TOPOLOGY: linear

133

(ii) MOLECULE TYPE: cDNA

(ix) FEATURES:

(B) LOCATION: 5, 7, 10, 13

5 (D) OTHER INFORMATION: G represents inosine (i)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

GGTCGTGCCG CAGGG

15

10

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:

15 (A) LENGTH: 21 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

TTAGGGATAG CCCTCATCTC T

21

25

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:

30 (A) LENGTH: 21 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

35 (ii) MOLECULE TYPE: cDNA

134

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

TCAGGGATAG CCCCCATCTA T

21

5

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 24 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

AACCCCTTGC CACTACATCA ATTT

24

20

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:

- 25 (A) LENGTH: 23 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

30 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

GCGTAAGGAC TCCTAGAGCT ATT

23

35

135

## (2) INFORMATION FOR SEQ ID NO: 25:

## (i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 18 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

TCATCCATGT ACCGAAGG

18

15

## (2) INFORMATION FOR SEQ ID NO: 26:

## (i) SEQUENCE CHARACTERISTICS:

- 20 (A) LENGTH: 20 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

25

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

ATGGGGTTCC CAAGTTCCT

20

## 30 (2) INFORMATION FOR SEQ ID NO: 27:

## (i) SEQUENCE CHARACTERISTICS:

- 35 (A) LENGTH: 20 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

136

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

5

GCCGATATCA CCCGCCATGG

20

(2) INFORMATION FOR SEQ ID NO: 28:

10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

15

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

20

GCATCCGGCA ACTGCACG

18

(2) INFORMATION FOR SEQ ID NO: 29:

25

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

30

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

35

CGCGATGCTG GTTGGAGAGC

20

137

## (2) INFORMATION FOR SEQ ID NO: 30:

- 5 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleotide
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

15 TCTCCACTCC GAATATTCCG

20

## (2) INFORMATION FOR SEQ ID NO: 31:

- 20 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 26 base pairs
  - (B) TYPE: nucleotide
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

25

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

30 GATCTAGGCC ACTTCTCAGG TCCAGS

26

## (2) INFORMATION FOR SEQ ID NO: 32:

- 35 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 23 base pairs

138

- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: cDNA

(ix) FEATURES:

- (B) LOCATION: 6, 12, 19
- (D) OTHER INFORMATION: G represents inosine (i)

10

(xi) SEQUENCE DESCRIPTION : SEQ ID NO: 32

CATCTGTTTG GGCAGGCAGT AGC

23

15

(2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: cDNA

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

CTTGAGCCAG TTTCATACC TGGA

24

30

(2) INFORMATION FOR SEQ ID NO: 34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single

35



139

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

AGTGYTRCCM CARGGCGCTG AA

22

10 (2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs

(B) TYPE: nucleotide

15 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

GMGGCCAGCA GSAKGCATC CA

22

(2) INFORMATION FOR SEQ ID NO: 36:

25

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

30 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

35

GGATGCCGCC TATAGCCTCT AC

22

140

## (2) INFORMATION FOR SEQ ID NO: 37:

- 5 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleotide
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

15 AAGCCTATCG CGTGCAGTTG CC

22

## (2) INFORMATION FOR SEQ ID NO: 38:

- 20 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 40 base pairs
  - (B) TYPE: nucleotide
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

25

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

30 TAAAGATCTA GAATTCGGCT ATAGGCGGCA TCCGGCAAGT

40

## (2) INFORMATION FOR SEQ ID NO: 39

- 35 (i) SEQUENCE CHARACTERISTICS :
- (A) LENGTH : 50 amino acids

141

(B) TYPE : amino acid

(ii) MOLECULE TYPE : peptide

5 (xi) SEQUENCE DESCRIPTION : SEQ ID NO: 39

Asp Ala Phe Phe Cys Ile Pro Val Arg Pro Asp Ser Gln Phe Leu Phe  
 1 5 10 15  
 Ala Phe Glu Asp Pro Leu Asn Pro Thr Ser Gln Leu Thr Trp Thr Val  
 10 20 25 30  
 Leu Pro Gln Gly Phe Arg Asp Ser Pro His Leu Phe Gly Gln Ala Leu  
 35 40 45  
 Ala Gln  
 50

15

(2) INFORMATION FOR SEQ ID NO: 40

(i) SEQUENCE CHARACTERISTICS :

20 (A) LENGTH : 150 base pairs  
 (B) TYPE : nucleic acid  
 (C) STRANDEDNESS : single  
 (D) TOPOLOGY : linear

(ii) MOLECULE TYPE : cDNA

25

(xi) SEQUENCE DESCRIPTION : SEQ ID NO: 40

GATGCCTTTT TCTGCATCCC TGTACGTCCT GACTCTCAAT TCTTGTTTGC CTTTGAAGAT 60  
 CCTTTGAACC CAACGTCTCA ACTCACCTGG ACTGTTTAC CCCAAGGGTT CAGGGATAGC 120  
 30 CCCCATCTAT TTGGCCAGGC ATTAGCCCAA 150

(2) INFORMATION FOR SEQ ID NO: 41

35 (i) SEQUENCE CHARACTERISTICS :

(A) LENGTH : 11 amino acids

142

(B) TYPE : amino acid

(ii) MOLECULE TYPE : peptide

5 (xi) SEQUENCE DESCRIPTION : SEQ ID NO: 41

Cys Ile Pro Val Arg Pro Asp Ser Gln Phe Leu

1 5 10

10 (2) INFORMATION FOR SEQ ID NO: 42

(i) SEQUENCE CHARACTERISTICS :

(A) LENGTH : 17 amino acids

(B) TYPE : amino acid

15

(ii) MOLECULE TYPE : peptide

(xi) SEQUENCE DESCRIPTION : SEQ ID NO: 42

20 Val Leu Pro Gln Gly Phe Arg Asp Ser Pro His Leu Phe Gly Glu Ala

1 5 10 15

Leu

17

25

(2) INFORMATION FOR SEQ ID NO: 43

(i) SEQUENCE CHARACTERISTICS :

(A) LENGTH : 8 amino acid

30 (B) TYPE : amino acid

(ii) MOLECULE TYPE : peptide

(xi) SEQUENCE DESCRIPTION : SEQ ID NO: 43

35

Leu Phe Ala Phe Glu Asp Pro Leu

143

1 5 8

(2) INFORMATION FOR SEQ ID NO: 44

5

(i) SEQUENCE CHARACTERISTICS :

(A) LENGTH : 8 amino acids

(B) TYPE : amino acid

10 (ii) MOLECULE TYPE : peptide

(xi) SEQUENCE DESCRIPTION : SEQ ID NO: 44

Phe Ala Phe Glu Asp Pro Leu Asn

15 1 5 8

(2) INFORMATION FOR SEQ ID NO: 45

20 (i) SEQUENCE CHARACTERISTICS :

(A) LENGTH : 25 base pairs

(B) TYPE : nucleic acid

(C) STRANDEDNESS : single

(D) TOPOLOGY : linear

25

(ii) MOLECULE TYPE : cDNA

(xi) SEQUENCE DESCRIPTION : SEQ ID NO: 45

30 GTGCTGATTG GTGTATTTAC AATCC

25

(2) INFORMATION FOR SEQ ID NO: 46

35 (i) SEQUENCE CHARACTERISTICS :

(A) LENGTH : 1859 base pairs

144

(B) TYPE : nucleic acid  
(C) STRANDEDNESS : single  
(D) TOPOLOGY : linear

5 (ii) MOLECULE TYPE : cDNA

(xi) SEQUENCE DESCRIPTION : SEQ ID NO: 46

```
GTGCTGATTG GTGTATTTAC AATCCTTTAT CTAATCCGAA ATGCCCATGT TGCAATATGG 60
10 AAAGAAAGGG AGTTCCTAAC CTCTGGGGGA ACCCCCATTA AATACCACAA GTAAATCATG 120
GAGTTATTGC ACACAGTGCA AAAACTCAAG GAGGTGGAAG TCTTACTG CCAAAGCCAT 180
CAGAAAAGGG AAGAGGGGAG AAGAGCAGCA TAAGTGGCTA CAGAGGCAAG GAAAGACTAG 240
CAGAAAGGAA AGAGAGAAAG AGACAGAAAG TCAGAGAGAG AGAGAGGAAG AGACAGAGCA 300
CAAAGAGGGA GTCAGAGAGA GAGAGAGACA GAGAGTCAGA GAGAAGGAAA GAGAGAGAGG 360
15 AAGAGACAAA GAATGAATCA AACAGAGAGA CAGAAAGTCA GAGAGAGAGA GAGAGAGGAA 420
GAGACAGAGA AAAAGAGGGA GTCAGAAAAA GAGAGACCAA AGAAGAAGTC CAAAGAGAAA 480
GAAAGAGAGA TGGAAAGTAGT AAAGGAAAAA CAGTGTACCC TATTCCTTTA AAAGCCGGGG 540
TAAATTTAAA ACCTATAATT GATAACTGAA GGTCTTCTCT GTAACCTGT AACACTCCAA 600
TACCACCTTG TTGTCAAGTG TAAACAAGGG CGTAGCCCAA AAGCACTGAG GCCACTAACA 660
20 ACCCATAGCC TTCCTATCAA AATTCCTTAA CCCAGCAGGT TTCCTAACAG GGGATCTAAA 720
TCTTAATTAA TTACCATACA ATGGTCCAAC CAGACTTAGG AGGAATTCCC TTCAGGACGG 780
GAAGATAGAT GCTTCCTCCC AGGCGATTAA GGGAGAAAGA CACAATGGGT ATTCAGTAAG 840
TGCCAAGGGG AACACTTGTA GAAGCAAAGT TAGGAAAATT GCCAAATAAT TGGTTTGCTC 900
AAGAGTTGTT TGCACTCAGC CAAACCTTGA AGTACTTGCA GAATCAGAAA GGAGCCATCT 960
25 ATACCAATTC TAAGTTAATA TGGACTGAAG GAGGTTTTAT TAATACCAA GAGAAATTAA 1020
AATCCCAAAC TTATAAGGTT TTCAACCAA GTAAAGTTTG CTAAAAGTTA ACAGCGTAAC 1080
ATGTATTATC CTACTACCAC ACACTCTCAA AGGATTTCTC AGACAGTTTG CAAGAAATAA 1140
TGATATCTAT CCTTACTCTA CAATCCCAA TAGACTCTTT GGCAGCAGTG ACTCTCCAA 1200
ACCGTCAAGG CCTAGACCTC CTCACTGCTG AGAAAGGAGG ACTCTGCACC TTCTTAAGGG 1260
30 AAGAGTGTTG TCTTTACTT AACCAGTCAG GGATAGTATG AGATGCTGCC CGGCATTTAC 1320
AGAAAAAGGC TTCTGAAATC AGACAACGCC TTTCAAATTC CTATACCAAC CTCTGGAGTT 1380
GGGCAACATG GTTCTTCCC TTTCTATGTC CCATGGCTGC CATCTTGCTA TTA CTGCTT 1440
TTGGGCCCCTG TATTTTAAAC CTCCTTGTC AATTTGTTTC TTCTAGGATC GAGGCCATCA 1500
AGCTACAGAT GGTCTTACAA ATGGAACCCC AAATGAGCTC AACTATCAAC TTCTACTGAG 1560
35 GACCCCTAGA CCAACCCCTT GGCCCTTTCA CTGGCCTAAA GAGTTCCCCT CTGGAGGACA 1620
CTACCACTGC AGGGCCCCAT CTTTGCCCCT ATCCAGAAGG AAGTAGCTAG AGCAGTCATT 1680
```

145

GCCCAATTCC CAAGAGCAGC TGGGGTGTCC CGTTTAGAGT GGGGATTGAG AGGTGAAGCC 1740  
AGCTGGACTT CTGGGTCGGG TGGGGACTTG GAGAACTTTT GTGTCTAGCT AAAGGATTGT 1800  
AAATGCAACA ATCAGTGCTC TGTGTCTAGC TAAAGGATTG TAAATACACC AATCAGCAC 1859

5

(2) INFORMATION FOR SEQ ID NO: 47

(i) SEQUENCE CHARACTERISTICS :

- 10 (A) LENGTH : 23 base pairs  
(B) TYPE : nucleic acid  
(C) STRANDEDNESS : single  
(D) TOPOLOGY : linear

15 (ii) MOLECULE TYPE : cDNA

(xi) SEQUENCE DESCRIPTION : SEQ ID NO: 47

TGATGTGAAC GGCATACTCA CTG

23

20

(2) INFORMATION FOR SEQ ID NO: 48

(i) SEQUENCE CHARACTERISTICS :

- 25 (A) LENGTH : 24 base pairs  
(B) TYPE : nucleic acid  
(C) STRANDEDNESS : single  
(D) TOPOLOGY : linear

30 (ii) MOLECULE TYPE : cDNA

(xi) SEQUENCE DESCRIPTION : SEQ ID NO: 48

CCCAGAGGTT AGGAACTCCC TTTC

24

35

146

## (2) INFORMATION FOR SEQ ID NO: 49

## (i) SEQUENCE CHARACTERISTICS :

- 5 (A) LENGTH : 25 base pairs  
(B) TYPE : nucleic acid  
(C) STRANDEDNESS : single  
(D) TOPOLOGY : linear

## (ii) MOLECULE TYPE : cDNA

10

## (xi) SEQUENCE DESCRIPTION : SEQ ID NO: 49

GCTAAAGGAG ACTTGTGGTT GTCAG

25

15

## (2) INFORMATION FOR SEQ ID NO: 50:

## (i) SEQUENCE CHARACTERISTICS:

- 20 (A) LENGTH: 22 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

25

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

CAACATGGGC ATTTCGGATT AG

22

30

## (2) INFORMATION FOR SEQ ID NO: 51:

## (i) SEQUENCE CHARACTERISTICS:

- 35 (A) LENGTH: 400 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single



147

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

GGCTGCTAAA GGAGACTTGT GGTGTCAGA CAATCGCCTA CTTAGGTACC AGGCCTTATT 60  
ACTTGAGGGA CTGGTGCTTC AGATGCGCAC TTGTGCAGCT CTTAACCCAA ACTTATGCTG 120  
CCCAGAAGGA TCTTTTAGAG GTCCCCTTAG CCAACCCTGA CCTCAACCTA TATATATACT 180  
10 GATGGAAGTT CGTTTGTAGA AAAGGGATTA CAAAGGGNAG GATATNCCAT AGGTTAGTGA 240  
TAAAGCAGTA CTTGAAAGTA AGCCTCTTCC CCCCAGGGAC CAGCGCCCCC GTTAGCAGAA 300  
CTAGTGGCAC TGACCCCGAG CCTTAGAACT TGGAAAGGGA GGAGGATAAA TGTGTATACA 360  
GATAGCAAGT ATGCTTATCT AATCCGAAAT GCCCATGTTC 400

15

(2) INFORMATION FOR SEQ ID NO: 52:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2389 base pairs  
20 (B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

TCAGGGATAG CCCCCATCTA TTTGGTCAGG CACTGGCCCA AGATCTAGGG ACATGCCACT 60  
TTTAAGAGCC ATTTCTCAAG TCCAGGTACT CTGGTCCTTC GGTATGTGGA TGATTTACTT 120  
30 TTGGCTACCA GTTCAGTAGC CTCATGCCAG CAGGCTACTC TAGATCTCTT GAACTTTCTA 180  
GCTAATCAAG GGTACAAGGC ATCTAGGTTG AAGGCCCAGC TTTGCCTACA GCAGGTCAAA 240  
TATCTAGGCC TAATCTTAGC CAGAGGGACC AGGGCACTCA GCAAGGAACA AATACAGCCT 300  
ATACTGGCTT ATCCTCACCC TAAGACATTA AAACAGTTGC GGGGGTTCCT TGGAATCACT 360  
GGCTTTTTTG TGA CTATGGA TTCCAGATA CAGCAAGATT GGCAGGCCCC TCTATACTGT 420  
35 AATCAAGGAG ACTCACGAGG GCAAGTACTC ATCTAGTAGA ATGGGA ACTA GGGACAGAAA 480  
CAGCCTTCAA AACCTTAAAG CAGGCCCTAG TACAATCTCC AGCTTTAAGC CTTCCACAG 540

148

GACAAACTT CTCTTTATAC ATCACAGAGA GGGCAGAGAT AGCTCTTGGT GTCCTTATTC 600  
 AGACTCATGG GACTACCCCA CAACCAGTGG CACACCTAAG TAAGGAAATT GATGTAGTAG 660  
 CAAAAGGCTG GCCTCACTGT TTATGGGTAG CTGTGGTGGT GGCTGTCTTA GTGTCAGAAG 720  
 CTATCAAAAT AATACAAGGA AAGGATCTCA CTGTCTGGAC TACTCATGAT GTAATGGCAT 780  
 5 ACTAGGTGCC AAAAGAAGTT TATGGGTATC AGACAACCAC CTGCTTAGAT ACCAGGGACT 840  
 ACTCCTGGAG GATTGGGCTT CAAGTGCGTT TTTTGTGGCC TCAACCCTGC CACTTTTCCT 900  
 CCAGAGGATG GAGAGCCGCT TGAGCATGCT TGCCAACAGG TTGTAGGCCA GAATTATTCC 960  
 ACCCGAGATG ATCTCTTAGA GTACCCTTAG CTAATCCTGA CCTTAACCTA TATACCAATG 1020  
 GAAGTTCATT TGTGGAAAAC GGGATATGAA GGGCAGGTTA TGTCATAGTT AGTGATGTAA 1080  
 10 TCATACTTGC AAGTAAGCCT CTTACCCCAG GGGCCAGCAC TCAGTTAGCA GAACTAGTCA 1140  
 CACTTACCTT AACCTTAGAA CTGGGAAAGG GAAAAAGAAT AAATATGTAT ACAGATAGTA 1200  
 AGTATGCTTA TCTAATCCTA CATGCCCATG CTGCAATATG GAAGGAAAGG GAGTTCCTAA 1260  
 CCCCTGGGGG AACCCCCATT AAATACCACA AGGYAAATCA TGGAGTTATT GCACGCAGTG 1320  
 CAAAAACTCA AGGAGGTGGC AGTCTTACAC TGCCGAAGCY ATCAAAAAGG GGAAGGAGAG 1380  
 15 GGGAGAACAG CAGCATAAGT GGTGGCAGA GGCAGTGAAA GACCAGCAGA GAGAAGGAGA 1440  
 GAGACAACGT CAACGACAGA AGGAAAGAAG AGGAGGAGAC AGAGAGGAAG AGACAGAGAG 1500  
 ACAGTTAGTC CAAGAGAGAG ACAGAGAGAG GAAGAGACAG ACAGAAAGTC CAAGAGAGAA 1560  
 GGAAAGAGAG GAAGAGACCA AGGAGTCCNA GAGAGAGAAA GAGATAGAAG TAGTAAAGAA 1620  
 AAAACATTGT ACCCTATTCC TTTAAAAGCC GGGGTATATT TAAAACCTAT AATTGATAAT 1680  
 20 TGAGTTCTTG CACCCTCCTC CAGGGGATYG CTGGGAGGAA ACCCTCAACC GATATGTGAA 1740  
 AATTGTGGGT CGTCCCTATG TCTCAATTAC CAGCCAATAC CCCCTTGTTT TTAGTGTGAA 1800  
 CGAGGGTGTA GAGCGCAGAC AGGGAGACCT CTGACAATCC ATACCCTTCC TATCCAAAAT 1860  
 CCTTAACCCA GCAGGTTTTT TAAAAGGGGA TCTAAATCTT AATTAATTAC CATACAAAGG 1920  
 TCAAACCAGA TCTAGGAGGA ACTTCCTTCA GGACAGGATG ATAGATGGTT CCTCCCAGGC 1980  
 25 GATTAAAGAA AATAAAAAGA CACATGGGCA GCCAGTAAGT GATAAGGGAA CACTAGTAGA 2040  
 AGCAGTTAGG AGAAGTTGCC TAATAATTGG TCTACTCCAA ATGTGTGAGT TGTTGCGACT 2100  
 CAGCCCAAAT CTTAAAGTAC TTACAGAATT AGGGAGGAGC CATTTACACC AATTCTAAGT 2160  
 TAATATGGAC TGGATGAGGT TTTATTAATA GCGAAGGAGA ATTAAATCCT AAACCTNACAA 2220  
 GGTTTTCAAC TAAAGTAAAT TTTACTAAAA GCTAACAGTG TAACATGCAT TATCCTACTA 2280  
 30 CAACACACTC TCANAGGATT CCTCAGACAG TTTACAAGAA ATAACAAAAT CTATCTGGTA 2340  
 AGGATAGTAA CTACAATCCC AAATACATTC TTTGGCAGCA GTGACTCTC 2389

(2) INFORMATION FOR SEQ ID NO: 53:

35

(i) SEQUENCE CHARACTERISTICS:

149

- (A) LENGTH: 2448 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

5

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

```
10 TCAGGGATAG CCCCATCTA TTTGATCAGG CACTAGCCCA AGATCTAGGC CACTTCTGAA 60
   GTCCAGGCAT TCTAGTCCTT CAGTATGTGG ATGATTTACT TTTGGCTACC AGTTTGAAG 120
   CCTCATGCCA GCAGGCTACT TGAGATCTCT TGAACTTTCT AGCTAATCAA GGGTGTATGG 180
   CATCTAAATT GAAAGTCCAG CTCTGCCTAC AACAAGTCAA ATATCTAGGC CTAATCTTAG 240
   ATAGAAGAAC CAGGGCCCTC AGCAAGGAAT GAATAAAGCC TATGCTGGCT TATCGGCACC 300
15 CTAAGACATT AAAACAATTG TGGGGGTTCC TTGGAATCAC TGGCTTTTGC CGACTATGGA 360
   TCCCTGGATA GAGTGAGATA GCCAGGCCCC CTCTATTACT CTTATCAAGG AGACCCAGAG 420
   GGCAAATACT TATCTAGTAT TATGGGNACC AGAGGCAGAA AAAGCCTTCC AAACCTTAAA 480
   GGAGACCCTA GTACAAGCTC CAGCTTTAAG CCTTCCCACA GGACAAANCT TCTCTTTATA 540
   TGTCACAGAG AGAGCAGGAA TAGCTCCTGG AGTCCTTACT CAGACTTTTG GACGACCCCA 600
20 CGGCCAGTGG CRTACCTAAG TAAGGAAATT GATGTAGTAG CAAAAGGCTG GCCTCACTGT 660
   TTATGGGTAG TTGCGGCTGT GGCAGTCTTA CTGTCAAAGG CTATCAAAAT AATACAAGGA 720
   AAGGATTTCA CTATCTGGAC TACTCATGAG GAAAATGGCA TATTAGGTGC CAAAGGAAGT 780
   TTTTGGCTAT CAGACAACCA CCTGCTCAGA TTCCAGGCAC TACTGATTGA GAGACCAGTG 840
   CTTTAAATAT GTATGTGTGT GTGTGGCCCT CAACCCTGCC ACTGTTCTCC CAGAAGATGG 900
25 AGAACCAATG AAGCATTACT GTCAACAAAT TAGAGTCCAG AGTTATGCTG CCTGAGAGGA 960
   TCTCTTAGAA GTCCCCTTAG CTAATCCTGA CCTTAACCTA TATGCTGATG GAAGTTCACT 1020
   TGTGGAGAAT GGGATACGAA AAGCACATTA TGCCATAGTT AGTGAGGTAA CAGTACTTGA 1080
   AAGTAAGCCT ATTCCCCCAT GGACCAGAGC CCAGTTAGCA GAAGTAGTGG CACTTACCCA 1140
   AGCCTTAGAA CTAGGAAAGG GAAAAATAAT AAATGTGTAT ACAGATAGCA AGTATGCTTA 1200
30 TCTAATCCTA CATGCCCATG CTGCAGTATG GAAAGAAAGG GAGTTCCTAA CCTCTGGGGG 1260
   AACCCCCATT AAATACCACA AGGCAAATCA TGGAGTTATT GCATGTAGTG CAAAACCTCA 1320
   AGTAGGTGGC AGTTTTACAC TGCCTGAAGC TATGGGGAAG GAGAGAGGAG AACAGCAGCA 1380
   TAAGTGGCTA GCAGAGGCAG CGAAAGACTA GCAGAGAGGA GAGGTAGGGG AAAGACAGAA 1440
   AGTCAAAGAA AAGAAGTCAA AGACAGACAG AGAAAGAGAC AGAGGGAGCC AGAGAGAAAG 1500
35 AAAAGAGAGA ACGAAAGAGA CAGAATGTCA AAGAACAGAA GAGAGAGGCA GCGCCAGAAG 1560
   AGTTAAGAAA GTGAGAAAGA GAGATGGAAA TAGTAAAGAA AAAACAGTGT ACCCTATTCC 1620
```

150

TTTAAAAGCC AGGGTAAATT TAAAACGTAT AATTTTATAA TTGGAAGGTC TTCTCCATAA 1680  
CCCTATAACA TTAAATACC ACCTTGTTGT CAGTGTAAC AAGAGCATAG CCCAAAAGCA 1740  
CTGAGGCCAC TGACAACCCA TAGCCTTCCT ATCAAAAATC CTTAACTCTG CAGGTTTCCT 1800  
AACAGGGGAT CTAAATCTCA ACTAATCACC ATACAATGGT CCGACCAGAC CTAGGAGCGA 1860  
5 CTCCCCTCAG GACAGAAGGA TGGATGGTTC CTCCCAGGCC ATTAAGGGAA AGAGACACAA 1920  
TGGGTATTCA GTAAGTGATA AGGGAACCTCT TGTAAGCA GTTAGGAAGA TTGCCTAATA 1980  
TTTGGTCTGC TCAAATGTGC CAGCTGTTTG CACTCAGCTA AACCTTAAAT TACTTACAGA 2040  
ATTAGGAAGG AGCCATCTAT ACCAATTCTG AGTTAATATG AGCTGAACAA GTTCTTATTA 2100  
ATAGCAAAGA ATCATTGAAA TCTCAAACCTT GCAAAGTTTT CAACAAAAGT AAAGTTTGCT 2160  
10 GAAAGTTAGC AGTGTAACAT GTATTATCCT AACTTCTAAT CTTGTGGAAA TCAGACCCTA 2220  
TCAGTGCCCC TCAAAGCTGA AGTCCATCAG CATATGGCCA TACAACTAAT ACCCCTATTT 2280  
ATAGGGTTAG GAATGGCCAC TGCTACAGGA ATGGGAGTAA CAGGTTTATC TACTTCATTA 2340  
TCCTATTACC ACACACTCTT AAAGGATTTC TCAGACAGTT TACAAGAAAT AACAAAATCT 2400  
ATCCTTACTC TNTARTCCCA AATAGRTTCT TTGGCAGCAG TGA CTCTC 2448  
15

(2) INFORMATION FOR SEQ ID NO: 54:

(i) SEQUENCE CHARACTERISTICS:

- 20 (A) LENGTH: 21 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

CCTGAGTTCT TGCACTAACC C 21  
30

(2) INFORMATION FOR SEQ ID NO: 55:

(i) SEQUENCE CHARACTERISTICS:

- 35 (A) LENGTH: 23 base pairs  
(B) TYPE: nucleotide

151

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

GTCCGTTGGG TTCCTTACT CCT

23

10

(2) INFORMATION FOR SEQ ID NO: 56:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1196 base pairs

15

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

	TTCCTGAGTT	CTTGCACTAA	CCTCAAATGA	GAGAAGTGCC	GCCATAACTG	CAACCCAAGA	60
	GTTTGGCGAT	CCCTGGTATC	TCAGTCAGGT	CAATGACAGG	ATGACAACAG	AGGAAAGATA	120
25	ATGATTCCCC	ACAGGCCAGC	AGGCAGTTCC	CAGTGTAGAC	CCTCATTAGG	ACACAGAATC	180
	AGAACATGGA	GATTGGTGCC	GCAGACATTT	GCTAACTTGC	GTGCTAGAAG	GACTAAGGAA	240
	AACTAGGAAG	ATATGAATTA	TTCAATGATG	TCCACTATAA	CACAGGGGAA	AGGAAGAAAA	300
	TCCTACTGCC	TTTCTGGAGA	GACTAAGGGA	GGCATTGAGG	AAGCATACCA	GGCAAGTGGA	360
	CATTGGAGGC	TCTGGAAAAG	GGAAAAGTTG	GGAAAAGTAT	ATGTCTAATA	GGGCTTGCTT	420
30	CCAGTGTGGT	CTACAAGGAC	ACTTTAAAAA	AGATTGTCCA	ATAGAAATAA	GCCACCACCT	480
	CGTCCATGCC	CCTTATGTCA	AGGGAATCAC	TGGAAGGCCC	ACTGCCCCAG	GGGATGAAGG	540
	TCCTCTGAGT	CAGAAGCCAC	TAACCAGATG	ATCCAGCAGC	AGGACTGAGG	GTGCCCAGGG	600
	CAAGCGCCAG	CCCATGCCAT	CACCCTCACA	GAGCCCCAGG	TATGCTTGAC	CATTGAGGGT	660
	CAGAAGGGTA	CTGTCTCCTG	GACACTGGCG	GGCCTTCTCA	GTCTTACTTT	CCTGTCTCTG	720
35	ACAAGTGTCC	TCCAGATCTG	TCACTGTCCG	AGGGGTCTTA	GGACAGCCAG	TCACTAGATA	780
	CTTCTCCCAG	CCACTAAGTT	GTGACTGGGG	AACTTTACTC	TTCCACATGC	TTTTCTAATT	840

152

ATGCCTGAAA GCCCACTCT CTTGTTAGGG GAGAGACATT CTAGCAAAG CAGGGGCCAT 900  
TATACATGTG AATATAGGAG AAGGAACAAC TGTTTGTGTG CCCCTGCTTG AGGAAGGAAT 960  
TAATCCTGAA GTCCGGGCAA CAGAAGGACA ATATGGACAA GCAAAGAATG CCCGTCCTGT 1020  
TCAAGTTAAA CTAAAGGATT CCACCTCCTT TCCCTACCAA AGGCAGTACC CCCTCAGACC 1080  
5 CGAGACCCAA CAAGAACTCC AAAAGATTGT AAAGGACCTA AAAGCCCAAG GCCTAGTAAA 1140  
ACCAAGCAAT AGCCCTTGCA AGACTCCAAT TTTAGGAGTA AGGAAACCCA ACGGAC 1196

(2) INFORMATION FOR SEQ ID NO: 57:

10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2391 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

15

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

20

ATGATCCAGC AGCAGGACNG AGGGTGCCCG GGGCAAGCGC CAGCCCATGC CATCACCCCTC 60  
ACAGAGCCCC AGGTATGCTT GACCATTGAG GGTGAGAAGG GTNACTGTCT CCTGGACACT 120  
GGCGGNGCCT TCTCAGTCTT ACTTTCCTGT CCTGGACAAC TGTCCTCCAG ATCTGTCACT 180  
GTCCGAGGGG TCCTAGGACA GCCAGTCACT AGATACTTCT CCCAGCCACT AAGTTGTGAC 240  
25 TGGGGAAGTT TACTCTTCCC ACATGCTTTT CTAATTATGC CTGAAAGCCC CACTCTCTTG 300  
TTGGGGAGAG ACATTCTAGC AAAAGCAGGG GCCATTATAC ATGTGAATAT AGGAGAAGGA 360  
ACAACTGTTT GTTGTCCCCT GCTTGAGGAA GGAATTAATC CTGAAGTCCG GGCAACAGAA 420  
GGACAATATG GACAAGCAAA GAATGCCCGT CCTGTTCAAG TTAAACTAAA GGATTCCACC 480  
TCCTTTCCCT ACCAAAGGCA GTACCCCTC AGACCCGAGA CCCAACAAGA ACTCCAAAAG 540  
30 ATTGTAAGG ACCTAAAGC CCAAGGCCTA GTAAACCAA GCAATAGCCC TTGCAAGACT 600  
CCAATTTTAG GAGTAAGGAA ACCCAACGGA CAGTGGAGGT TAGTGCAAGA ACTCAGGATT 660  
ATCAATGAGG CTGTTGTTCC TCTATACCCA GCTGTACCTA ACCCTTATAC AGTGCTTTCC 720  
CAAATACCAG AGGAAGCAGA GTGGTTTACA GTCCTGGACC TTAAGGATGC CTTTTCTGC 780  
ATCCCTGTAC GTCCTGACTC TCAATTCTTG TTTGCCTTTG AAGATCCTTT GAACCCAACG 840  
35 TCTCAACTCA CCTGGACTGT TTTACCCCAA GGGTTTCAGG ATAGCCCCCA TCTATTGGC 900  
CAGGCATTAG CCCAAGACTT GAGTCAATTC TCATACCTGG AACTCTTGT CCTTCAGTAC 960

153

ATGGATGATT TACTTTTAGT CGCCCGTTCA GAAACCTTGT GCCATCAAGC CACCCAAGAA 1020  
CTCTTAACTT TCCTCACTAC CTGTGGCTAC AAGGTTTCCA AACCAAAGGC TCGGCTCTGC 1080  
TCACAGGAGA TTAGATACTN AGGGCTAAAA TTATCCAAAG GCACCAGGGC CCTCAGTGAG 1140  
GAACGTATCC AGCCTATACT GGCTTATCCT CATCCCCAAA CCCTAAAGCA ACTAAGAGGG 1200  
5 TTCCTTGGCA TAACAGGTTT CTGCCGAAAA CAGATTCCCA GGTACASCCC AATAGCCAGA 1260  
CCATTATATA CACTAATTAN GGAAACTCAG AAAGCCAATA CCTATTTAGT AAGATGGACA 1320  
CCTACAGAAG TGGCTTTCCA GGCCCTAAAG AAGGCCCTAA CCCAAGCCCC AGTGTTTCAGC 1380  
TTGCCAACAG GGCAAGATTT TTCTTTATAT GCCACAGAAA AAACAGGAAT AGCTCTAGGA 1440  
GTCCTTACGC AGGTCTCAGG GATGAGCTTG CAACCCGTGG TATACCTGAG TAAGGAAATT 1500  
10 GATGTAGTGG CAAAGGGTTG GCCTCATNGT TTATGGGTAA TGGNGGCAGT AGCAGTCTNA 1560  
GTATCTGAAG CAGTTAAAAT AATACAGGGA AGAGATCTTN CTGTGTGGAC ATCTCATGAT 1620  
GTGAACGGCA TACTCACTGC TAAAGGAGAC TTGTGGTTGT CAGACAACCA TTTACTTAAN 1680  
TATCAGGCTC TATTACTTGA AGAGCCAGTG CTGNGACTGC GCACTTGTGC AACTCTTAAA 1740  
CCCCAACTTA TGCTGCCCAG AAGGATCTTT NTAGAGGTCC CCTTAGCCAA CCCTGACCTC 1800  
15 AACTATATAT ATACTGATGG AAGTTCGTTT GTAGAAAAGG GATTACAAAG GGNAGGATAT 1860  
NCCATAGGTG TTAGTGATAA AGCAGTACTT GAAAGTAAGC CTCTTCCCCC CCAGGGACCA 1920  
GCGCCCCCGT TAGCAGAACT AGTGGCACTG ACCCCGCGAG CCTTAGAACT TTGGAAAGGG 1980  
AGGAGGATAA ATGTGTATAC AGATAGCAAG TATGCTTATC TAATCCGAAA TGCCCATGTT 2040  
GTTTATCTAA TCCGAAATGC CCATGTTGCA ATATGGAAAG AAAGGGAGTT CCTAACCTCT 2100  
20 GGGGGAACCC CCATTAAATA CCACAAGTTA ATCATGGAGT TATTGCACAC AGTGCAAAAA 2160  
CTCAAGGAGG TGGAAGTCTT ACACTGCCAA AGCCATCAGA AAAGGGAAAAG GGGAGAAGAG 2220  
CAGCATAAGT GGCTACAGAG GCAAGGAAAAG ACTAGCAGAA AGGAAAGAGA GAAAGAGACA 2280  
GAAAGTCAGA GAGAGAGAGA GGAAGAGACA GAGCACAAAG AGGGAGTCAG AGAGAGAGAG 2340  
AGACAGAGAG TCAGAGAGAA GGAAAGAGAG AGAGGAAGAG ACAAAGAATG A 2391  
25

(2) INFORMATION FOR SEQ ID NO: 58:

(i) SEQUENCE CHARACTERISTICS:

- 30 (A) LENGTH: 1722 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

35 (ii) MOLECULE TYPE: cDNA

154

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

```

TGGAGAATAG CAGCATAAGT TGGCTGGCAG AAGTAGGGAA AGACAGCAAG AAGTAAAGAA   60
AAAAARGAGA AAGTCAGAGA AAGAAAAAAA GAGAGGAAGA AACAAAGAAG AACTTGAAGA   120
5  GAGAAAGAAG TAGTAAAGAA AAAACAGTAT ACCCTATTCC TTTAAAAGCC AGGGTAAATT   180
TCTGTCTACC TAGCCAAGGC ATATTCTTCT TATGTGGAAC ATCAACCTAT ATCTGCCTCC   240
CCACTAACTG GACAGGCACC TGAACCTTAG TCTTTCTAAG TCCCAACATT AACATTGCCC   300
CAGGAAATCA GACCCTATTG GTACCTGTCA AAGCTAAAGT CCCGTCAGTG CAGAGCCATA   360
CAACTAATAT CCCTATTTAT AGGGTTAGGA ATGGCTACTG CTACAGGAAC TGGAAATAGCC   420
10 GGTTTATCTA CTTCAATTAT CTACTACCAT AACTCTCAA AGAATTTCTC AGACAGTTTG   480
CAAGAAATAA TGAAATCTAT TCTTACTTTA CAATCCCAAT TAGACTCTTT GGCAGCAATG   540
ACTCTCCAAA ACCGCCGAGG CCCACACCTC CTCACTGCTG AGAAAGGAGG ACTCTGCACC   600
TTCTTAGGGG AAGAGTGTTG TTTTACACT AACCAGTCAG GGATAGTACG AGATGCCACC   660
TGGCATTTAC AGGAAAGGGC TTCTGATATC AGACAATGCC TTTCAAACCTC TTATACCAAC   720
15 CTCTGGAGTT GGGCAACATG GCTTCTTCCA TTTCTAGGTC CCATGGCAGC CATCTTGCTG   780
TTACTCACCT TTGGGCCCTG TATTTTAAAG CTTCTTGTC AATTTGTTTC CTCTAGGATC   840
GAAGCCATCA AGCTACAGAT GGTCTTACAA ATGGAACCCC AAATGAGTTC AACTAACAAC   900
TTCTACCAAG GACCCCTGGA ACGATCCACT GGCACCTCCA CTAGCCTAGA GATTCCCCCTC   960
TGAAGACAC TACAACTGCA GGGCCCCTTC TTTGCCCTA TCCAGCAGGA AGTAGCTAGA 1020
20 GCGGTCATCG GCCAAATTCC CAACAGCAGT TGGGGTGTC TGTTTAGAGG GGGGATTGAA 1080
GAGGTGACAG CCTGCTGGCA GCCTCACAGC CCTCGTTGGY TCTCAGTGCC TCCTCAGCCT 1140
TGGTGCCAC TCTGGCCGTG CTTGAGGAGC CCTTCAGCCT GCCACTGCAC TGTGGGAGCC 1200
TCTTTCTGGG CTGGACAAGG CCGGAGCCAG CTCCTCAGC TTGCAGGGAG GTATGGAGGG 1260
AGAGATGCAG GCGGGAACCA GGGCTGCGCA TGGCGCTTGC GGGCCAGCAT GAGTTCCAGG 1320
25 TGGGCGTGGG CTCGGCGGGC CCCACACTCG GGCAGTGAGG GGCTTAGCAC CTGGGCCAGA 1380
CAGATGCTGT GCTCAACTTC TTCGCTGGGC CTTAGCTGCC TTCCCCGTGG GGCAGGGCTY 1440
CGGGAACMTG CAGCCTGCCC ATGCTTGAGC CCCCCACCCC GCCGTGGGTT CYTGCACAGC 1500
CCAAGCTTCC CGGACAAGCA CCACCCCTTA TCCACGGTGC CCAGTCCCAT CAACCACCCA 1560
AGGGTTGAGG AGTGCGGGCA CACAGCGCGG GATTGGCAGG CAGTTCCACT TGCGGCCTTG 1620
30 GTGCGGGATC CACTGCGTGA AGCCAGCTGG GCTCCTGAGT CTGGTGGGGA CTTGGAGAAT 1680
CTTTATGTCT AGCTAAGGGA TTGTAAATAC ACCAATCAGC AC                               1722

```

(2) INFORMATION FOR SEQ ID NO: 59:

35

(i) SEQUENCE CHARACTERISTICS:



155

- (A) LENGTH: 495 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

```
10  CTTCCCCAAC TAATAAGGAC CCCCTTTTCA ACCCAAACAG TCCAAAAGGA CATAGACAAA 60
    GGAGTAAACA ATGAACCAA GAGTGCCAAT ATTCCCTGGT TATGCACCCT CCAAGCGGTG 120
    GGAGAAGAAT TCGGCCCAGC CAGAGTGCAT GTACCTTTTT CTCTCTCACA CTTGAAGCAA 180
    ATTAAATAG ACNTAGGTNA ATTNTCAGAT AGCCCTGATG GYTATATTGA TGTTTTACAA 240
    GGATTAGGAC AATCCTTTGA TCTGACATGG AGAGATATAA TATTACTGCT AAATCAGACG 300
15  CTAACCTCAA ATGAGAGAAG TGCTGCCATA ACTGGAGCCC GAGAGTTTGG CAATCTCTGG 360
    TATCTCAGTC AGGTCAATGA TAGGATGACA ACGGAGGAAA GAGAACGATT CCCCACAGGG 420
    CAGCAGGCAG TTCCCAGTGT AGCTCCTCAT TGGGACACAG AATCAGAACA TGGAGATTGG 480
    TGCCGCAGAC ATTTA                                     495
```

20

(2) INFORMATION FOR SEQ ID NO: 60:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2503 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

25

(ii) MOLECULE TYPE: cDNA

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

```
35  CCAAGAACCC ACCAATTCCG GANACATTT TGCGGACCAC GAAGGGACTT TCGCATATCG 60
    CCAAGCGGTG AGACAATAGC CGAGCGGTGA GACCTTTCCC AATCGCCAAG CAGTGAGTAC 120
    CATCAGACCC CTTTCACTTG CTATTCTGTC CTATCTTTCT TTAGAATTCT GGGGCTAAAT 180
    ACCGGGCATC TGTCAGCCAT TTAAAAGTGA CTAGCGGGCC GCCGGACTAA AGACACGGGT 240
```

156

GTCAAGCTTT CTGGGAAAGG GCTCTCTAAC AACCCCCAAC TCTTTGGAGT TGGGACCGTT 300  
 GGTTTGCCTA GAACCAGCTT CCGCTTTTCC TGTACTTCTG GGCTGAGCCG TGGGTTGACA 360  
 GTGAAGGAAA GCCATGCATC TCCGGGGTCT CGMCAACATG TTGGTTGACC CTGCGGCCAT 420  
 GAGTGGAACT CTCAAAAGCA TGTGCCCCAA GCGACACTCG CCTATCTATC CTATCTATCC 480  
 5 TGACCCCTTGC CCTCTGGGTC CTAATGCCTG CCAGACAAAC TTCCTCTCGC CTCTCTTCTC 540  
 TGAAGCTAGA ACCGCTTCTA AAAATTGCTA CCTGGTCTCT GGTGCTTTTC CTARTTTCTC 600  
 CTATAAAGAA TGAWTTCTAG TATTAAACTC CAGGACTCTG TTACCTTCTT TAGGCACCCG 660  
 GGCTCACCAA TCAGAAAGAC ACAGTTTTTG CCCAAGGCCC CATCGTAGTG GGGACTACCT 720  
 GGAATTTTAG GATCCCTCCT CAGACTAACA GGCCTAACAA AAGTTATTCC TGAAGCTAGG 780  
 10 ATATGGGGAG CCTCAGAAAT TGTATCCCTC CTATTCATAT AAGTGAGAAC AAAAGGTGTC 840  
 ACTCTTCCAA CCCTGAAGAT CCCCTCCCTC CCTCAGGGTA TGGCCCTCCA TTTCATTTTT 900  
 GTGGCATAAC ATCTTTATAG GATGGGGTAA AGTCCCAATA CTAACAGGAG AATGCTTAGG 960  
 ACTCTAACAG GTTTTTGAGA ATGCGTCAGT AAGGGCCACT AAATCTGATT TTTCTCAGTC 1020  
 GGTCCCTCCT GTGGTCTAGG AGGACAGGCA AGGTTGTGCA GGTTTTCGAG AATGCGTCAG 1080  
 15 TAAGGACCAC TAAATCCGAC CTTCCTCGGT CCTCCATGTG GTCTGGGAGG AAAACTAGTG 1140  
 TTTCTGCTGC TGCCTCGGTG AGCGCAACTA TTCAAGTCAG CAGGGTCCAG GGACCGTTGC 1200  
 AGGTTCTTGG GCAGGGGTTG TTTCTGCTGC TGCATTGGTG AATGCAACTA TTCTGATCAG 1260  
 CAGGGTCCCA GGACCATTCG AGGTCCTTGG GCAGGGAGAG AAACAAAACA AACCAAAACT 1320  
 GTGGGCGGTT TTGTCTTTCA TATGGGAAAC ACTCAGGCAT CAACAGGTTC ACCCTTGAAA 1380  
 20 TGCATCCTAA GCCATTGGGA CCAATTTGAC CCACAAACCC TGAAAAGAG GAGGCTCATT 1440  
 TTTTCCTGCA CTACGGCTTG GCCCCAATAT TCTCTTTYTG ATGGGGAAAA ATGGCCACCT 1500  
 GAGGGAAGCA CAAATTACAA TAYTATCCTA CAGCYTGATC TTTTCTGTAA GAGGGAAGGC 1560  
 AAATGGAGTG AATACCTTAT GTCCAAGCTT TCTTTTCATT GAGGGAGAAT ACACAACTAT 1620  
 GCAAAGCTTG CAATTTACAT CCCACAGGAG GACCCTTCAG CTTACCCCCA TATCCTAGCC 1680  
 25 TCCCTATAGC TTCCCTTCCT ATTGATGATA CTCCTCCTCT AATCTCCCCT GCCCAGAAGG 1740  
 AAATAAGCAA AGAAATCTCC AAAGGTCCAC AAAAACCCCC GGGCTATCGG TTATGTCCCT 1800  
 TCAAGYTGTA GGGGGAGGGG AATTTGGCCC AACCCGGGTG CATGTCCCTT CTCCCTCTCT 1860  
 GATTTAAAGC AGATCAAGGC AGACCTGGGG AAGTTTTTCAG ATGATCCTGA TAGGTACATA 1920  
 GATGTCCTAC AGGGTCTAGG GCAAACCTTT GACCTCACTT GGAGAGACGT CATGCTACTG 1980  
 30 TTAGATCAAA CCCTGGCCTT TAATGAAAAG AATGCGGCTT TAGCTGCAGC CTGAGAGTTT 2040  
 GGAGATACCT GGTATCCTAG TCAAGTAAAT GAAAGAATGA CAGCCGAAGA AAGGGACAAC 2100  
 TTCCTTACTG GTCAGCAACC CATCCCCAGT ATGGATCCCC ACTGGGACTT TGA CTCAGAT 2160  
 CATGGGGACT GGAGTCGTAA ACATCTGTTG ATCTGTGTTT TGGAAGGACT AAGGAGAATT 2220  
 GGGAAAAAGC CCATGAATTA TTCAATGATA TCCACCATAA CCCAGGGAAA GGAAGAAAAT 2280  
 35 CCTTCTGCCT TCCTCGAGCG GCTACAAGAG GCCTTAAGAA AATATACTCC CCTGTCACCC 2340  
 GAATCACTCG AGGGTCAATT GATTCTAAAA GATAAGTTTA TTACCCAATC AGCCACAGAT 2400

157

ATCAGGAGAA AGCTCCAAAA GCAAGCCCTG AGCCTGAACA AAATCTAGAG ACATTATTAA 2460  
 ACCTGGCAAC CTTGGTGTTT TATAATAGGG ACCAAGAGGA ACA 2503

## 5 (2) INFORMATION FOR SEQ ID NO: 61:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1167 base pairs  
 (B) TYPE: nucleotide  
 10 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## 15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

AAGGAACTC AGAAAGCCAA TACCCATTTA GTAAGATGGA CACCAGAAGC AGAAGCAGCT 60  
 TTCCAGGCCC TAAAGAAATC CCTAACCCAA GCCCCAGTGT TAAGCTTGCC AACGGGGCAA 120  
 GACTTTTCTT TATATGTCAC AGAAAAACAG GAATAGCTCT AGGAGTCCTT ACACAGGTCC 180  
 20 AAGGGACAAG CTTGCAACCT GTGGCATAAC TGAGTAAGGA AACTGATGTA NTGGCAAAGG 240  
 GTTGGCCTCA TTGTTTACAG GTAGGGCAGC AGTAGCAGTC TTAGTTTCTG AAACAGTTAA 300  
 AATAATACAG GGAAGAGATC TTAGTGTGTG GACATCTCAT GATGTGAACG GCATACTCAC 360  
 TGCTAAAGAG GACTTGTGGC TGTCAGACAA CCATTTACTT AAATAGCAGG TTCTATTACT 420  
 TGAAGTGCCA GTGCTGCGAC TGCACATTTG TGCAACTCTT AACCAGCCA CATTCTTCC 480  
 25 AGACAATGAA GAAAAGATAG AACATAACTG TCAACAAGTA ATTGCTCAA CCTATGCTGC 540  
 TCGAGGGGAC CTTCTAGAGG TTCCCTTGAC TGATCCCGAC CTCAACTGT ATACTGATGG 600  
 AAGTTCCTTG GCAGAAAAAG GACTTTGAAA AGCGGGGTAT GCAGTGATCA GTGATAATGG 660  
 AATACTTGAA AGTAATCGCC TCACTCCAGG AACTAGTGCT CACCTGGCAG AACTAATAGC 720  
 CCTCACTTGG GCACTAGAAT TAGGAGAAGG AAAAAGGGTA AATATATATT CAGACTCTAA 780  
 30 GTATGCTTAC CTAGTCCTCC ATGCCCATGC AGCAATATGG AGAGAGAGGG AATTCCTAAC 840  
 TTCTGAGGGA ACACCTATCA ACCATCAGGG AAGCCATTAG GAGATTATTA TTGGCTGTAC 900  
 AGAAACCTAA AGAGGTGGCA GTCTTACACT GCCAGGGTCA TCAGGAAGAA GAGGAAAGGG 960  
 AAATAGAAGG CAATCGCCAA GCGGATATTG AAGCAAAAAA AGCCGCAAGG CAGGACTCTC 1020  
 CATTAGAAAT GCTTATAGAA GGACCCCTAG TATGGGGTAA TCCCCTCTGG GAAACCAAGC 1080  
 35 CCCAGTACTC AGCAGGAAAA ATAGAATAGG AAACCTCACA AGGACATACT TTCCTCCCCT 1140  
 CCAGATGGCT AGCCACTGAG GAAGGAA 1167

158

## (2) INFORMATION FOR SEQ ID NO: 62:

- 5 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 78 base pairs
  - (B) TYPE: nucleotide
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

15 TCCAAAGGCA CCAGGGCCCT CAGTGAGGAA CGTATCCAGC CTATACTGGC TTATCCTCAT 60  
 CCCAAAACCC TAAAGCAA 78

## (2) INFORMATION FOR SEQ ID NO: 63

20

- (i) SEQUENCE CHARACTERISTICS :
- (A) LENGTH : 26 amino acids
  - (B) TYPE : amino acid

25 (ii) MOLECULE TYPE : peptide

(xi) SEQUENCE DESCRIPTION : SEQ ID NO: 63

Ser Lys Gly Thr Arg Ala Leu Ser Glu Glu Arg Ile Gln Pro Ile Leu  
 30 1 5 10 15  
 Ala Tyr Pro His Pro Lys Thr Leu Lys Gln  
 20 25

## 35 (2) INFORMATION FOR SEQ ID NO: 64:

159

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 28 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

5

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

10

AAATGTCTGC GGCACCAATC TCCATGTT

28

## (2) INFORMATION FOR SEQ ID NO: 65:

15

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

20

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

25

AAGGGGCATG GACGAGGTGG TGGCTTATTT

30

## (2) INFORMATION FOR SEQ ID NO: 66:

30

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

35

(D) TOPOLOGY: linear

160

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

GGAGAAGAGC AGCATAAGTG G

21

5

(2) INFORMATION FOR SEQ ID NO: 67:

(i) SEQUENCE CHARACTERISTICS:

10

(A) LENGTH: 25 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

15

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

GTGCTGATTG GTGTATTTAC AATCC

25

20

(2) INFORMATION FOR SEQ ID NO: 68:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 34 base pairs

25

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

GACTCGCTGC AGATCGATTT TTTTTTTTTT TTTT

34

35 (2) INFORMATION FOR SEQ ID NO: 69:

161

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

5 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

10

GCCATCAAGC CACCCAAGAA CTCTTAAGT

30

## (2) INFORMATION FOR SEQ ID NO: 70:

15

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

20 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

25 CCAATAGCCA GACCATTATA TACACTAATT

30

## (2) INFORMATION FOR SEQ ID NO: 71:

## 30 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

35 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

162

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

GCCATAACTG CAACCCAAGA GTT

23

5

(2) INFORMATION FOR SEQ ID NO: 72:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 base pairs

10

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

GGACGAGGTG GTGGCTTATT TCT

23

20

(2) INFORMATION FOR SEQ ID NO: 73:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs

25

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

AACTTGCGTG CTAGAAGGAC TAAGG

25

35

(2) INFORMATION FOR SEQ ID NO: 74:



163

## (i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 24 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## 10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

AACTTTTCCC TTTCCAGAT CCTC

24

## 15 (2) INFORMATION FOR SEQ ID NO: 75:

## (i) SEQUENCE CHARACTERISTICS:

- 20 (A) LENGTH: 22 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

GCATACCAGG CAAGTGGACA TT

22

## 30 (2) INFORMATION FOR SEQ ID NO: 76:

## (i) SEQUENCE CHARACTERISTICS:

- 35 (A) LENGTH: 25 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

164

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

5

CTGTCCGTTG GGTTCCTTA CTCCT

25

(2) INFORMATION FOR SEQ ID NO: 77:

10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

15

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 77:

20

GAGGCTCTGG AAAAGGGAAA AGTT

24

(2) INFORMATION FOR SEQ ID NO: 78:

25

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

30

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

35

CTGTCCGTTG GGTTCCTTA CTCCT

25

165

## (2) INFORMATION FOR SEQ ID NO: 79:

- 5 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 25 base pairs
  - (B) TYPE: nucleotide
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

15 AGGAGTAAGG AAACCCAACG GACAG

25

## (2) INFORMATION FOR SEQ ID NO: 80:

- 20 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 25 base pairs
  - (B) TYPE: nucleotide
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

TGTATATAAT GGTCTGGCTA TTGGG

25

30

## (2) INFORMATION FOR SEQ ID NO: 81:

- 35 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 25 base pairs
  - (B) TYPE: nucleotide

166

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

AGGAGTAAGG AAACCCAACG GACAG

25

10

(2) INFORMATION FOR SEQ ID NO: 82:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 base pairs

15

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

TTCGGCAGAA ACCTGTTATG CCAAGG

26

25

(2) INFORMATION FOR SEQ ID NO: 83:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs

30

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

167

CTCGATTCTTCT TGCTGGGCCT TA

22

## 5 (2) INFORMATION FOR SEQ ID NO: 84:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleotide
- 10 (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## 15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

GTTGATTCCC TCCTCAAGCA

20

## 20 (2) INFORMATION FOR SEQ ID NO: 85:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleotide
- 25 (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## 30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

CTCTACCAAT CAGCATGTGG

20

## 35 (2) INFORMATION FOR SEQ ID NO: 86:

168

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

5 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

10

TGTTCTCTT GGTCCCTAT

19

## (2) INFORMATION FOR SEQ ID NO: 87:

15

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 433 aminoacids

(B) TYPE: aminoacid

20

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 87:

Met Ala Thr Ala Thr Gly Thr Gly Ile Ala Gly Leu Ser Thr Ser Leu

1

5

10

15

25

Ser Tyr Tyr His Thr Leu Ser Lys Asn Phe Ser Asp Ser Leu Gln Glu

20

25

30

Ile Met Lys Ser Ile Leu Thr Leu Gln Ser Gln Leu Asp Ser Leu Ala

35

40

45

Ala Met Thr Leu Gln Asn Arg Arg Gly Pro His Leu Leu Thr Ala Glu

30

50

55

60

Lys Gly Gly Leu Cys Thr Phe Leu Gly Glu Glu Cys Cys Phe Tyr Thr

65

70

75

80

Asn Gln Ser Gly Ile Val Arg Asp Ala Thr Trp His Leu Gln Glu Arg

85

90

95

35

Ala Ser Asp Ile Arg Gln Cys Leu Ser Asn Ser Tyr Thr Asn Leu Trp

100

105

110

169

	Ser Trp Ala Thr Trp Leu Leu Pro Phe Leu Gly Pro Met Ala Ala Ile	
	115	120 125
	Leu Leu Leu Leu Thr Phe Gly Pro Cys Ile Phe Lys Leu Leu Val Lys	
	130	135 140
5	Phe Val Ser Ser Arg Ile Glu Ala Ile Lys Leu Gln Met Val Leu Gln	
	145	150 155 160
	Met Glu Pro Gln Met Ser Ser Thr Asn Asn Phe Tyr Gln Gly Pro Leu	
		165 170 175
	Glu Arg Ser Thr Gly Thr Ser Thr Ser Leu Glu Ile Pro Leu Trp Lys	
10		180 185 190
	Thr Leu Gln Leu Gln Gly Pro Phe Phe Ala Pro Ile Gln Gln Glu Val	
	195	200 205
	Ala Arg Ala Val Ile Gly Gln Ile Pro Asn Ser Ser Trp Gly Val Leu	
	210	215 220
15	Phe Arg Gly Gly Ile Glu Glu Val Thr Ala Cys Trp Gln Pro His Ser	
	225	230 235 240
	Pro Arg Trp Xaa Ser Val Pro Pro Gln Pro Trp Cys Pro Leu Trp Pro	
		245 250 255
	Cys Leu Arg Ser Pro Ser Ala Cys His Cys Thr Val Gly Ala Ser Phe	
20		260 265 270
	Trp Ala Gly Gln Gly Arg Ser Gln Leu Pro Gln Leu Ala Gly Arg Tyr	
	275	280 285
	Gly Gly Arg Asp Ala Gly Gly Asn Gln Gly Cys Ala Trp Arg Leu Arg	
	290	295 300
25	Ala Ser Met Ser Ser Arg Trp Ala Trp Ala Arg Arg Ala Pro His Ser	
	305	310 315 320
	Gly Ser Glu Gly Leu Ser Thr Trp Ala Arg Gln Met Leu Cys Ser Thr	
		325 330 335
	Ser Ser Leu Gly Leu Ser Cys Leu Pro Arg Gly Ala Gly Leu Arg Glu	
30		340 345 350
	Xaa Ala Ala Cys Pro Cys Leu Ser Pro Pro Pro Arg Arg Gly Phe Leu	
	355	360 365
	His Ser Pro Ser Phe Pro Asp Lys His His Pro Leu Ser Thr Val Pro	
	370	375 380
35	Ser Pro Ile Asn His Pro Arg Val Glu Glu Cys Gly His Thr Ala Arg	
	385	390 395 400

170

Asp Trp Gln Ala Val Pro Leu Ala Ala Leu Val Arg Asp Pro Leu Arg  
405 410 415  
Glu Ala Ser Trp Ala Pro Glu Ser Gly Gly Asp Leu Glu Asn Leu Tyr  
420 425 430  
5 Val  
433

(2) INFORMATION FOR SEQ ID NO: 88:

10

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 693 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
15 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 88:

20

CTTCCCCAAC TAATAAGGAC CCCCCTTTCA ACCCAAACAG TCCAAAAGGA CATAGACAAA 60  
GGAGTAAACA ATGAACCAAA GAGTGCCAAT ATTCCCTGGT TATGCACCCT CCAAGCGGTG 120  
GGAGAAGAAT TCGGCCAGC CAGAGTGCAT GTACCTTTTT CTCTCTCACA CTTGAAGCAA 180  
ATTAAAATAG ACNTAGGTNA ATTNTCAGAT AGCCCTGATG GYTATATTGA TGTTTTACAA 240  
25 GGATTAGGAC AATCCTTTGA TCTGACATGG AGAGATATAA TATTACTGCT AAATCAGACG 300  
CTAACCTCAA ATGAGAGAAG TGCTGCCATA ACTGGAGCCC GAGAGTTTGG CAATCTCTGG 360  
TATCTCAGTC AGGTCAATGA TAGGATGACA ACGGAGGAAA GAGAACGATT CCCACAGGG 420  
CAGCAGGCAG TTCCAGTGT AGCTCCTCAT TGGGACACAG AATCAGAACA TGGAGATTGG 480  
TGCCGCAGAC ATTTACTAAC TTGCGTGCTA GAAGGACTAA GGAAACTAG GAAGACTATG 540  
30 AATTATTCAA TGATGTCCAC TATAACACAG GGGAAAGGAA GAAAATCCTA CTGCCTTTCT 600  
GGAGAGACTA AGGGAGGCAT TGAGGAAGCA TACCAGGCAA GTGGACATTG GAGGCTCTGG 660  
AAAAGGGAAA AGTTGGGCAA ATTGAATGCC TAA 693

35 (2) INFORMATION FOR SEQ ID NO: 89:



171

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1577 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

5

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 89:

10

```
AACTTGCGTG CTAGAAGGAC TAAGGAAAAC TAGGAAGACT ATGAATTATT CAATGATGTC 60
CACTATAACA CAGGGGAAAG GAAGAAAATC CTAAGGGAGG 120
CATTGAGGAA GCATACCAGG CAAGTGGACA TTGGAGGCTC TGGAAAAGGG AAAAGTTGGG 180
CAAATTGAAT GCCTAATAGG GCTTGCTTCC AGTGCAGTCT ACAAGGACGC TTTAGAAAAG 240
15 ATTGTCCAAG TAGAAATAAG CCGCCCCTCG TCCATGCCCC TTATGTCAAG GGAATCACTG 300
GAAGGCCTAC TGCCCCAGGG GACGAAGGTC CTCTGAGTCA GAAGCCACTA ACCTGATGAT 360
CCAGCAGCAG GACTGAGGGT GCGCGGGGCA AGTGCCAGCC CATGCCATCA CCCTCAGAGC 420
CCCGGGTATG TTTGACCATT GAGAGCCAGG AAGTTAACTG TCTCCTGGAC ACTGGCGCAG 480
CCTTCTCAGT CTTACTTTCC TGTCCCAGAC AATTGTCCTC CAGATCTGTC ACTATCCGAG 540
20 GGGTCCTAAG ACAGCCAGTC ACTACATACT TCTCTCAGCC ACTAAGTTGT GACTGGGGAA 600
CTTTACTCTT TTCACATGCT TTTCTAATTA TGCCTGAAAG CCCCCTCCC TTGTTAGGGA 660
GAGACATTTT AGCAAAAGCA GGGGCCATTA TACACCTGAA CATAGGAAAA GGAATACCCA 720
TTTGCTGTCC CCTGCTTGAG GAAGGAATTA ATCCTGAAGT CTGGGCAATA GAAGGACAAT 780
ATGGACAAGC AAAGAATGCC CGTCCTGTTC AAGTTAACT AAAGGATTCT GCCTCCTTTC 840
25 CCTACCAAAG GAAGTACCCT CTTAGACCCG AGGCCCTACA AGGACTCAAA AGATTGTTAA 900
GGACCTAAAA GCCCAAGGCC TAGTAAAACC ATGCAGTAGC CCCTGCAATA CTCCAATTTT 960
AGGAGTAAGG AAACCAACG GACAGTGGAG GTTAGTGCAA GATCTCAGGA TTATTAATGA 1020
GGCTGTTTTT CCTCTATACC CAGCTGTATC TAGCCCTTAT ACTCTGCTTT CCCTAATACC 1080
AGAGGAAGCA GAGTAGTTTA CAGTCCTGGA CCTTAAGGAT GCCTCTTTCT GCATCCCTGT 1140
30 ACATCCTGAT TCTCAATTCT TGTTTGTCTT TGAAGATCCT TTGAACCCAA TGTCTCAATT 1200
CACCTGGACT GTTTTACCCC AGGGGTTCG GGATAGCCCC CATCTATTTG GCCAGGCATT 1260
AGCCCAAGAC TTGAGCCAAT TCTCATACCT GGACATCTTG TCCTTCGGTA TGGGATGATT 1320
TAATTTTAGC CACCCGTTCA GAAACCTTGT GCCATCAAGC CACCCAAGCG TTCTTAAATT 1380
TCCTCACTCC GTGTGGCTAC AAGGTTTCCA AACCAAAGGC TCAGCTCTGC TCACAGCAGG 1440
35 TTAAATACTT AGGGTTAAAA TTATCCAAAG GCACCAGGGC CCTCTGTGAG GAATGTATCC 1500
AACCTGTACT GGCTTATCTT CATCCCAAAA CCCTAAAGCA ACTAAGAAGG TCCTTGGCAT 1560
```

172

AACAGGTTTC TGCCGAA

1577

(2) INFORMATION FOR SEQ ID NO: 90:

5

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 182 amino acids

(B) TYPE: amino acid

10 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 90:

15	Ser Ser Ser Arg Thr Glu Gly Ala Arg Gly Lys Cys Gln Pro Met Pro	1	5	10	15
	Ser Pro Ser Glu Pro Arg Val Cys Leu Thr Ile Glu Ser Gln Glu Val	20	25	30	
	Asn Cys Leu Leu Asp Thr Gly Ala Ala Phe Ser Val Leu Leu Ser Cys	35	40	45	
20	Pro Arg Gln Leu Ser Ser Arg Ser Val Thr Ile Arg Gly Val Leu Arg	50	55	60	
	Gln Pro Val Thr Thr Tyr Phe Ser Gln Pro Leu Ser Cys Asp Trp Gly	65	70	75	80
	Thr Leu Leu Phe Ser His Ala Phe Leu Ile Met Pro Glu Ser Pro Thr	85	90	95	
25	Pro Leu Leu Gly Arg Asp Ile Leu Ala Lys Ala Gly Ala Ile Ile His	100	105	110	
	Leu Asn Ile Gly Lys Gly Ile Pro Ile Cys Cys Pro Leu Leu Glu Glu	115	120	125	
30	Gly Ile Asn Pro Glu Val Trp Ala Ile Glu Gly Gln Tyr Gly Gln Ala	130	135	140	
	Lys Asn Ala Arg Pro Val Gln Val Lys Leu Lys Asp Ser Ala Ser Phe	145	150	155	160
	Pro Tyr Gln Arg Lys Tyr Pro Leu Arg Pro Glu Ala Leu Gln Gly Leu	165	170	175	
35	Lys Arg Leu Leu Arg Thr				

173

180

## (2) INFORMATION FOR SEQ ID NO: 91:

5

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 36 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

10

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 91:

15

AGATCTGCAG AATTCGATAT CACCCCCCCC CCCCCC

36

## (2) INFORMATION FOR SEQ ID NO: 92:

20

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

25

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 92:

30

AGATCTGCAG AATTCGATAT CA

22

## (2) INFORMATION FOR SEQ ID NO: 93:

## (i) SEQUENCE CHARACTERISTICS:

35

(A) LENGTH: 2304 base pairs

(B) TYPE: nucleotide

174

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 93:

5	TCCAGCAGCA GGACTGAGGG TGCCCGGGGC AAGTGCCAGC CCATGCCATC	50
	ACCCTCAGAG CCCCAGGTAT GTTTGACCAT TGAGAGCCAG GAAGTTAACT	100
	GTCTCCTGGA CACTGGCGCA GCCTTCTCAG TCTTACTTTC CTGTCCCAGA	150
	CAATTGTCCT CCAGATCTGT CACTATCCGA GGGGTCCTAG GACAGCCAGT	200
	CACTACATAC TTCTCTCAGC CACTAAGTTG TGAAGGGGA ACTTTACTCT	250
10	TTTCACATGC TTTTCTAATT ATGCCTGAAA GCCCCACTCC CTTGTTAGGG	300
	AGAGACATTT TAGCAAAAGC AGGGGCCATT ATACACCTGA ACATAGGAAA	350
	AGGAATACCC ATTTGCTGTC CCCTGCTTGA GGAAGGAATT AATCCTGAAG	400
	TCTGGGCAAT AGAAGGACAA TATGGACAAG CAAAGAATGC CCGTCCTGTT	450
	CAAGTTAAAC TAAAGGATTC TGCCTCCTTT CCCTACCAA GGAAGTACCC	500
15	TCTTAGACCC GAGGCCCTAC AAGGANCTCA AAAGATTGTT AAGGACCTAA	550
	AAGCCCAAGG CCTAGTAAAA CCATGCAGTA GCCCCTGCAA TACTCCAATT	600
	TTAGGAGTAA GGAAACCCAA CGGACAGTGG AGGTTAGTGC AAGATCTCAG	650
	GATTATTAAT GAGGCTGTTT TTCCTCTATA CCCAGCTGTA TCTAGCCCTT	700
	ATACTCTGCT TTCCCTAATA CCAGAGGAAG CAGAGTGGTT TACAGTCCTG	750
20	GACCTTAAGG ATGCCTTTTT CTGCATCCCT GTACGTCCTG ACTCTCAATT	800
	CTTGTTTGCC TTTGAAGATC CTTTGAACCC AACGTCTCAA CTCACCTGGA	850
	CTGTTTTACC CCAAGGGTTC AGGGATAGCC CCCATCTATT TGGCCAGGCA	900
	TTAGCCCAAG ACTTGAGTCA ATTCTCATAC CTGGACACTC TTGTCCTTCA	950
	GTACGTGGAT GATTTACTTT TAGTCGCCCG TTCAGAAACC TTGTGCCATC	1000
25	AAGCCACCCA AGAACTCTTA ACTTTCCTCA CTACCTGTGG CTACAAGGTT	1050
	TCCAAACCAA AGGCTCGGCT CTGCTCACAG GAGATTAGAT ACTTAGGGCT	1100
	AAAATTATCC AAAGGCACCA GGGCCCTCAG TGAGGAACGT ATCCAGCCTA	1150
	TACTGGCTTA TCCTCATCCC AAAACCCTAA AGCAACTAAG AGGGTTCCTT	1200
	GGCATAACAG GTTTCTGCCG AAAACAGATT CCCAGGTACA CCCCATAAGC	1250
30	CAGACCATTA TATACATAA TTAGGGAAAC TCAGAAAGCC AATACCTATT	1300
	TAGTAAGATG GACACCTACA GAAGTGGCTT TCCAGGCCCT AAAGAAGGCC	1350
	CTAACCCAAG CCCAGTGTT CAGCTTGCCA ACAGGGCAAG ATTTTCTTT	1400
	ATATGCCACA GAAAAACAG GAATAGCTCT AGGAGTCCTT ACGCAGGTCT	1450
	CAGGGATGAG CTTGCAACCC GTGGTATACC TGAGTAAGGA AATTGATGTA	1500
35	GTGGCAAAGG GTTGGCCTCA TTGTTTATGG GTAATGGCGG CAGTAGCAGT	1550
	CTTAGTATCT GAAGCAGTTA AAATAATACA GGGAAGAGAT CTTACTGTGT	1600

175

	GGACATCTCA	TGATGTGAAC	GGCATACTCA	CTGCTAAAGG	AGACTTGTGG	1650
	TTGTCAGACA	ACCATTTACT	TAATTATCAG	GCTCTATTAC	TTGAAGAGCC	1700
	AGTGCTGAGA	CTGCGCACTT	GTGCAACTCT	TAAACCCGCC	ACATTTCTTC	1750
	CAGACAATGA	AGAAAAGATA	GAACATAACT	GTCAACAAGT	AATTGCTCAA	1800
5	ACCTATGCTG	CTCGAGGGGA	CCTTCTAGAG	GTTCCCTTGA	CTGATCCCGA	1850
	CCTCAACTTG	TATACTGATG	GAAGTTCCTT	GGCAGAAAAA	GGACTTCGAA	1900
	AAGCGGGGTA	TGCAGTGATC	AGTGATAATG	GAATACTTGA	AAGTAATCGC	1950
	CTCACTCCAG	GAAGTAGTGC	TCACCTGGCA	GAAGTAATAG	CCCTCACTTG	2000
	GGCACTAGAA	TTAGGAGAAG	GAAAAAGGGT	AAATATATAT	TCAGACTCTA	2050
10	AGTATGCTTA	CCTAGTCCTC	CATGCCCATG	CAGCAATATG	GAGAGAGAGG	2100
	GAATTCCTAA	CTTCTGAGGG	AACACCTATC	AACCATCAGG	AAGCCATTAG	2150
	GAGATTATTA	TTGGCTGTAC	AGAAACCTAA	AGAGGTGGCA	GTCTTACACT	2200
	GCCAGGGTCA	TCAGGAAGAA	GAGGAAAGGG	AAATAGAAGG	CAATCGCCAA	2250
	GCGGATATTG	AAGCAAAAAA	AGCCGCAAGG	CAGGACTCTC	CATTAGAAAT	2300
15	GCTT					2304

## (2) INFORMATION FOR SEQ ID NO: 94:

## (i) SEQUENCE CHARACTERISTICS:

- |    |                             |
|----|-----------------------------|
|    | (A) LENGTH: 2364 base pairs |
| 20 | (B) TYPE: nucleotide        |
|    | (C) STRANDEDNESS: single    |
|    | (D) TOPOLOGY: linear        |

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 94:

25	ATGATCCAGC	AGCAGGACNG	AGGGTGCCCCG	GGGCAAGCGC	CAGCCCATGC	50
	CATCACCCCTC	ACAGAGCCCC	AGGTATGCTT	GACCATTGAG	GGTCAGAAGG	100
	GTNACTGTCT	CCTGGACACT	GGCGGNGCCT	TCTCAGTCTT	ACTTTCCTGT	150
	CCTGGACAAC	TGTCCTCCAG	ATCTGTCACT	GTCCGAGGGG	TCCTAGGACA	200
	GCCAGTCACT	AGATACTTCT	CCCAGCCACT	AAGTTGTGAC	TGGGGAACCT	250
30	TACTCTTCCC	ACATGCTTTT	CTAATTATGC	CTGAAAGCCC	CACTCTCTTG	300
	TTGGGGAGAG	ACATTCTAGC	AAAAGCAGGG	GCCATTATAC	ATGTGAATAT	350
	AGGAGAAGGA	ACAACTGTTT	GTTGTCCCCT	GCTTGAGGAA	GGAATTAATC	400
	CTGAAGTCCG	GGCAACAGAA	GGACAATATG	GACAAGCAAA	GAATGCCCGT	450
	CCTGTTCAAG	TTAAACTAAA	GGATTCCACC	TCCTTTCCCT	ACCAAAGGCA	500
35	GTACCCCTC	AGACCCGAGA	CCCAACAAGA	ACTCCAAAAG	ATTGTAAAGG	550
	ACCTAAAAGC	CCAAGGCCTA	GTAAAACCAA	GCAATAGCCC	TTGCAAGACT	600

176

	CCAATTTTAG	GAGTAAGGAA	ACCCAACGGA	CAGTGGAGGT	TAGTGCAAGA	650
	ACTCAGGATT	ATCAATGAGG	CTGTTGTTCC	TCTATACCCA	GCTGTACCTA	700
	ACCCTTATAC	AGTGCTTTCC	CAAATACCAG	AGGAAGCAGA	GTGGTTTACA	750
	GTCCTGGACC	TTAAGGATGC	CTTTTTCTGC	ATCCCTGTAC	GTCCTGACTC	800
5	TCAATTCTTG	TTTGCCTTTG	AAGATCCTTT	GAACCCAACG	TCTCAACTCA	850
	CCTGGACTGT	TTTACCCCAA	GGGTTCAAGG	ATAGCCCCCA	TCTATTTGGC	900
	CAGGCATTAG	CCCAAGACTT	GAGTCAATTC	TCATACCTGG	ACACTCTTGT	950
	CCTTCAGTAC	ATGGATGATT	TACTTTTAGT	CGCCCGTTCA	GAAACCTTGT	1000
	GCCATCAAGC	CACCCAAGAA	CTCTTAACCT	TCCTCACTAC	CTGTGGCTAC	1050
10	AAGGTTTCCA	AACCAAAGGC	TCGGCTCTGC	TCACAGGAGA	TTAGATACTN	1100
	AGGGCTAAAA	TTATCCAAAG	GCACCAGGGC	CCTCAGTGAG	GAACGTATCC	1150
	AGCCTATACT	GGCTTATCCT	CATCCCAAAA	CCCTAAAGCA	ACTAAGAGGG	1200
	TTCCTTGGA	TAACAGGTTT	CTGCCGAAAA	CAGATTCCCA	GGTACASCCC	1250
	AATAGCCAGA	CCATTATATA	CACTAATTAN	GGAAACTCAG	AAAGCCAATA	1300
15	CCTATTTAGT	AAGATGGACA	CCTACAGAAG	TGGCTTTCCA	GGCCCTAAAG	1350
	AAGGCCCTAA	CCCAAGCCCC	AGTGTTCAAG	TTGCCAACAG	GGCAAGATTT	1400
	TTCTTTATAT	GCCACAGAAA	AAACAGGAAT	AGCTCTAGGA	GTCCTTACGC	1450
	AGGTCTCAGG	GATGAGCTTG	CAACCCGTGG	TATACCTGAG	TAAGGAAATT	1500
	GATGTAAGTG	CAAAGGGTTG	GCCTCATNGT	TTATGGGTAA	TGGNGGCAGT	1550
20	AGCAGTCTNA	GTATCTGAAG	CAGTTAAAT	AATACAGGGA	AGAGATCTTN	1600
	CTGTGTGGAC	ATCTCATGAT	GTGAACGGCA	TACTSRCTGC	TAAAGGAGAC	1650
	TTGTGGTTGT	CAGACAACCA	TTTACTTAAN	TAYCAGGCYY	TATTACTTGA	1700
	AGAGCCAGTG	CTGNACTGTC	GCACTTGTC	AACTCTTAAA	CCCAAACCTA	1750
	TGCTGCCCAG	AAGGATCTTT	NTAGAGGTCC	CCTTAGCCAA	CCCTGACCTC	1800
25	AACTATATAT	ATACTGATGG	AAGTTCGTTT	GTAGAAAAGG	GATTACAAAG	1850
	GGNAGGATAT	NCCATAGGTG	TTAGTGATAA	AGCAGTACTT	GAAAGTAAGC	1900
	CTCTTCCCCC	CCAGGGACCA	GCGCCCCCGT	TAGCAGAACT	AGTGGCACTG	1950
	ACCCCGCGAG	CCTTAGAACT	TTGGAAAAGG	AGGAGGATAA	ATGTGTATAC	2000
	AGATAGCAAG	TATGCTTATC	TAATCCGAAA	TGCCCATGTT	GCAATATGGA	2050
30	AAGAAAGGGA	GTTCCCTAAC	TCTGGGGGAA	CCCCCATTA	ATACCACAAG	2100
	TTAATCATGG	AGTTATTGCA	CACAGTGCAA	AAACTCAAGG	AGGTGGAAGT	2150
	CTTACACTGC	CAAAGCCATC	AGAAAAGGGA	AAGAGGGGAA	GAGCAGCATA	2200
	AGTGGCTACA	GAGGCAAGGA	AAGACTAGCA	GAAAGGAAAG	AGAGAAAGAG	2250
	ACAGAAAGTC	AGAGAGAGAG	AGAGGAAGAG	ACAGAGCACA	AAGAGGGAGT	2300
35	CAGAGAGAGA	GAGAGACAGA	GAGTCAGAGA	GAAGGAAAGA	GAGAGAGGAA	2350
	GAGACAAAGA	ATGAH				2365

177

## (2) INFORMATION FOR SEQ ID NO: 95:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 768 amino acids

5

(B) TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 95:

	SSSRTEGARG KCQPMSPSE PRVCLTIESQ EVNCLLDTGA AFSVLLSCPR	50
	QLSSRSVTIR GVLGQPVTTY FSQPLSCDWG TLLFSHAFLI MPESPTPLLG	100
10	RDILAKAGAI IHLNIGKGIP ICCPLLEEGI NPEVWAIEGQ YGQAKNARPV	150
	QVKLKDSASF PYQRKYPLRP EALQGQXQKIV KDLKAQGLVK PCSSPCNTPI	200
	LGVRKPNGQW RLVQDLRIIN EAVFPLYPAV SSPYTLLSLI PEEAEWFTVL	250
	DLKDAFFCIP VRPDSQFLFA FEDPLNPTSQ LTWTVLPQGF RDSPHLFGQA	300
	LAQDLSQFSY LDTLVLYVD DLLLVARSET LCHQATQELL TFLTTCGYKV	350
15	SKPKARLCSQ EIRYLGLKLS KGTRALSEER IQPILAYPHP KTLKQLRGFL	400
	GITGFCRKQI PRYTPRIARPL YTLIRETQKA NTYLVRWTPT EVAFAQALKKA	450
	LTQAPVFSLP TGQDFSLYAT EKTGIALGVL TQVSGMSLQP VVYLSKEIDV	500
	VAKGWPHCLW VMAAVAVLVS EAVKIIQGRD LTVWTSHDVN GILTAKGDLW	550
	LSDNHLNLYQ ALLLEEPVLR LRTCATLKPA TFLPDNEEKI EHNCQQVIAQ	600
20	TYAARGDLE VPLTDPDLNL YTDGSSSLAEK GLRKAGYAVI SDNGILESNR	650
	LTPGTSAPHLA ELIALTWALE LGEGKRVNIY SDSKYAYLVL HAHAAIWRER	700
	EFLTSEGTPV NHQEAIRRLI LAVQKPKEVA VLHCQGHQEE EEREIEGNRQ	750
	ADIEAKKAAR QDSPLEML	768

## 25 (2) INFORMATION FOR SEQ ID NO: 96:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 114 amino acids

(B) TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 96:

30

	SSSRTEGARG KCQPMSPSE PRVCLTIESQ EVNCLLDTGA AFSVLLSCPR	50
	QLSSRSVTIR GVLGQPVTTY FSQPLSCDWG TLLFSHAFLI MPESPTPLLG	100
	RDILAKAGAI IHLN	114

## 35 (2) INFORMATION FOR SEQ ID NO: 97:

## (i) SEQUENCE CHARACTERISTICS:

178

(A) LENGTH: amino acids

(B) TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 97:

5 IGKGIPICCPLEEINPEVWAI EGQY GQAKNARPV  
QVKLKDSASF PYQRKYPLRPEALQGXQKIVKDLKAQGLVKPCSSPCNTPI  
LGVRKPNGQWRLVQDLRIINEAVFPLYPAVSSPYTLLSLIPEEA EWFTVL  
DLKDAFFCIPVRPDSQFLFAFEDPLNPTSQLTWTVLPQGFRDSPHLFGQA  
LAQDLSQFSYLDTLVLQYVDDLLVARSETLCHQATQELLTFLTTCGYKV  
10 SKPKARLCSQEIRYLGLKLSKGTRALSEERIQPILAYPHPKTLKQLRGFL  
GITGFCRKQIPRYTPIARPLYTLIRETQKANTYLVRWTPTEVAFQALKKA  
LTQAPVFSLPTGQDFSLYATEKTGIALGVLTQVSGMSLQPVVYLSKEIDV  
VAKGWPHCLWVMAAVVLVSEAVKIIQGRDLTVWTS HDVNGI LTAKGDLW  
LSDNHLLNYQALLLEEPVLRRLRTCATLKPATFLPDNEEKIEHNCQQVIAQ  
15 TYAARGDLLLEVPLTDPDLNLYTDGSSSLAEKGLRKAGYAVISDNGILES NR  
LTPG TSAHLAELIALTWALELGEGKRVNIYSDSKYAYLVLHAHA AIWRER  
EFLTSEGTPINHQEAI RRLLLAVQKPKEVAVLHCQGHQEEEEEREIEGNRQ  
ADIEAKKAARQDSPLEML

20

(2) INFORMATION FOR SEQ ID NO: 98:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: amino acids

(B) TYPE: peptide

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 98:

LYTDGSSSLAEKGLRKAGYAVISDNGILES NR  
LTPG TSAHLAELIALTWALELGEGKRVNIYSDSKYAYLVLHAHA AIWRER  
EFLTSEGTPINHQEAI RRLLLAVQKPKEVAVLHCQGHQEEEEEREIEGNRQ  
30 ADIEAKKAARQDSPLEML

(2) INFORMATION FOR SEQ ID NO: 99

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 base pairs

35

(B) TYPE: nucleotide

(C) STRANDEDNESS: single



179

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 99:

AGGAGTAAGG AAACCCAACG GAC

23

5 (2) INFORMATION FOR SEQ ID NO: 100

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

10 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 100:

TAAGAGTTGC ACAAGTGCG

19

(2) INFORMATION FOR SEQ ID NO: 101

15 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 101:

TCAGGGATAG CCCCCATCTA T

21

(2) INFORMATION FOR SEQ ID NO: 102

(i) SEQUENCE CHARACTERISTICS:

25 (A) LENGTH: 24 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 102:

30 AACCCCTTGC CACTACATCA ATTT

24

(2) INFORMATION FOR SEQ ID NO: 103

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs

35 (B) TYPE: nucleotide

(C) STRANDEDNESS: single

180

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 103:

AGCAGCAGGA CTGAGGGT

18

5 (2) INFORMATION FOR SEQ ID NO: 104

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

10 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 104:

CTGTCCGTTG GGTTTCCTTA CTCCT

25

(2) INFORMATION FOR SEQ ID NO: 105

15 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 105:

GACAGCAAAT GGGTATTCCT TTCC

24

(2) INFORMATION FOR SEQ ID NO: 106

(i) SEQUENCE CHARACTERISTICS:

25 (A) LENGTH: 24 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 106:

30 AGGAGTAAGG AAACCAACG GACA

24

(2) INFORMATION FOR SEQ ID NO: 107

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs

35 (B) TYPE: nucleotide

(C) STRANDEDNESS: single

181

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 107:

TGATATAAT GGTCTGGCTA TTGGG

25

## 5 (2) INFORMATION FOR SEQ ID NO: 108

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

10 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 108:

TTCGGCAGAA ACCTGTTATG CCAAGG

26

## (2) INFORMATION FOR SEQ ID NO: 109

15 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 109:

GGCTCTGCTC ACAGGAGATT AGATAC

26

## (2) INFORMATION FOR SEQ ID NO: 110

(i) SEQUENCE CHARACTERISTICS:

25 (A) LENGTH: 26 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 110:

30 AAAGGCACCA GGGCCCTCAG TGAGGA

26

## (2) INFORMATION FOR SEQ ID NO: 111

(i) SEQUENCE CHARACTERISTICS:

35 (A) LENGTH: base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

182

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 111:

GGTTTAAGAG TTGCACAAGT GCGCAGTC

28

5 (2) INFORMATION FOR SEQ ID NO: 112:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 310 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

10 (D) TOPOLOGY: linear

(ii) TYPE DE MOLECULE: ADNc

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 112:

GCTTATAGAA GGACCCCTAG TATGGGGTAA TCCCTCTGG GAAACCAAGC CCCAGTACTC 60  
 AGCAGGAAAA ATAGAATAGG AAACCTCACA AGGACATACT TTCCTCCCCT CCAGATGGCT 120  
 15 AGCCACTGAG GAAGGAAAAA TACTTTTACC TGCAGCTAAC CAACAGAAAT TACTTAAAAC 180  
 CCTTCACCAA ACCTTCCACT TAGGCATTGA TAGCACCCAT CAGATGGCCA AATTATTATT 240  
 TACTGGACCA GGCCTTTTCA AAACATATCAA GAAGATAGTC AGGGGCTGTG AAGTGTGCCA 300  
 AAGAAATAAT 310

20 (2) INFORMATION FOR SEQ ID NO: 113:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 103 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

25 (D) TOPOLOGY: linear

(ii) TYPE DE MOLECULE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 113:

Leu Ile Glu Gly Pro Leu Val Trp Gly Asn Pro Leu Trp Glu Thr Lys  
 1 5 10 15  
 30 Pro Gln Tyr Ser Ala Gly Lys Ile Glu Xaa Glu Thr Ser Gln Gly His  
 20 25 30  
 Thr Phe Leu Pro Ser Arg Trp Leu Ala Thr Glu Glu Gly Lys Ile Leu  
 35 40 45  
 Ser Pro Ala Ala Asn Gln Gln Lys Leu Leu Lys Thr Leu His Gln Thr  
 35 50 55 60  
 Phe His Leu Gly Ile Asp Ser Thr His Gln Met Ala Lys Leu Leu Phe

183

65                                      70                                      75                                      80  
 Thr Gly Pro Gly Leu Phe Lys Thr Ile Lys Lys Ile Val Arg Gly Cys  
                                     85                                      90                                      95  
 Glu Val Cys Gln Arg Asn Asn  
 5                                      100

## (2) INFORMATION FOR SEQ ID NO: 114:

## (i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 635 base pairs  
 (B) TYPE: nucleotide  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) TYPE DE MOLECULE: ADNc

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 114:

15 CCCTGTATCT TTAACCTCCT TGTTAAGTTT GTCTCTTCCA GAATCAAAAC TGTA AAAACTA 60  
 CAAATTGTTT TTCAAATGGA GCACCAGATG GAGTCCATGA CTAAGATCCA CCGTGGACCC 120  
 CTGGACCGGC CTGCTAGCCC ATGCTCCGAT GTTAATGACA TTGAAGGCAC CCCTCCCGAG 180  
 GAAATCTCAA CTGCACAACC CCTACTATGC CCCAATTGAG CCGGAAGCAG TTAGAGCGGT 240  
 CATCAGCCAA CCTCCCCAAC AGCACTTGGG TTTTCCTGTT GAGAGGGGGG ACTGAGAGAC 300  
 20 AGGACTAGCT GGATTTCTTA GGCCAACGAA GAATCCCTAA GCCTAGCTGG GAAGGTGACT 360  
 GCATCCACCT CTAAACATGG GGCTTGCAAC TTAGCTCACA CCCGACCAAT CAGAGAGCTC 420  
 ACTAAAATGC TAATTAGGCA AAAATAGGAG GTAAAGAAAT AGCCAATCAT CTATTGCCTG 480  
 AGAGCACAGC GGGAGGGACA AGGATCGGGA TATAAACCCA GGCATTCGAG CCGGCAACGG 540  
 CAACCCCTT TGGGTCCCCT CCCTTTGTAT GGGCGCTCTG TTTTCACTCT ATTTCACTCT 600  
 25 ATTAAATCTT GCAACTGAAA AAAAAAAAAA AAAAA 635

## (2) INFORMATION FOR SEQ ID NO: 115:

## (i) SEQUENCE CHARACTERISTICS:

- 30 (A) LENGTH: 77 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) TYPE DE MOLECULE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 115:

35 Pro Cys Ile Phe Asn Leu Leu Val Lys Phe Val Ser Ser Arg Ile Lys  
                     1                                      5                                      10                                      15

184

Thr Val Lys Leu Gln Ile Val Leu Gln Met Glu His Gln Met Glu Ser  
 20 25 30  
 Met Thr Lys Ile His Arg Gly Pro Leu Asp Arg Pro Ala Ser Pro Cys  
 35 40 45  
 5 Ser Asp Val Asn Asp Ile Glu Gly Thr Pro Pro Glu Glu Ile Ser Thr  
 50 55 60  
 Ala Gln Pro Leu Leu Cys Pro Asn Ser Ala Gly Ser Ser  
 65 70 75

## 10 (2) INFORMATION FOR SEQ ID NO: 116:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs  
 (B) TYPE: nucleotide  
 (C) STRANDEDNESS: single  
 15 (D) TOPOLOGY: linear

## (ii) TYPE DE MOLECULE: ADNc

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 116:

TGGGGTTCCA TTTGTAAGAC CATCTGTAGC TT

32

## 20 (2) INFORMATION FOR SEQ ID NO: 117:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1481 base pairs  
 (B) TYPE: nucleotide  
 (C) STRANDEDNESS: single  
 25 (D) TOPOLOGY: linear

## (ii) TYPE DE MOLECULE: ADNc

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 117:

ATGGCCCTCC CTTATCATAC TTTTCTCTTT ACTGTTCTCT TACCCCTTT CGCTCTCACT 60  
 GCACCCCTC CATGCTGCTG TACAACCACT AGCTCCCTT ACCAAGAGTT TCTATGAAGA 120  
 30 ACGCGGCTTC CTGGAAATAT TGATGCCCCA TCATATAGGA GTTTATCTAA GGGAACTCC 180  
 ACCTTCACTG CCCACACCCA TATGCCCCGC AACTGCTATA ACTCTGCCAC TCTTTGCATG 240  
 CATGCAAATA CTCATTATTG GACAGGGAAA ATGATTAATC CTAGTTGTCC TGGAGGACTT 300  
 GGAGCCACTG TCTGTTGGAC TTAATTCACT CATAACAGTA TGTCTGATGG GGGTGAATT 360  
 CAAGGTCAGG CAAGAGAAAA ACAAGTAAAG GAAGCAATCT CCCAACTGAC CCGGGGACAT 420  
 35 AGCACCCCTA GCCCTACAA AGGACTAGTT CTCTCAAAAC TACATGAAAC CCTCCGTACC 480  
 CATACTCGCC TGGTGAGCCT ATTTAATACC ACCCTCACTC GGCTCCATGA GGTCTCAGCC 540

185

CAAAACCCTA CTAAGTGTG GATGTGCCTC CCCCTGCACT TCAGGCCATA CATTTCAATC 600  
 CCTGTTTCCTG AACAAATGGAA CAACTTCAGC ACAGAAATAA ACACCACTTC CGTTTTAGTA 660  
 GGACCTCTTG TTTCCAATCT GGAAATAACC CATACTCAA ACCTCACCTG TGTAATAATT 720  
 AGCAATACTA TAGACACAAC CAGCTCCCAA TGCATCAGGT GGGTAACACC TCCCACACGA 780  
 5 ATAGTCTGCC TACCCTCAGG AATATTTTTT GTCTGTGGTA CCTCAGCCTA TCATTGTTTG 840  
 AATGGCTCTT CAGAATCTAT GTGCTTCCTC TCATTCTTAG TGCCCCCTAT GACCATCTAC 900  
 ACTGAACAAG ATTTATACAA TCATGTCGTA CCTAAGCCCC ACAACAAAAG AGTACCCATT 960  
 CTTCTTTTTG TTATCAGAGC AGGAGTGCTA GGCAGACTAG GTACTGGCAT TGGCAGTATC 1020  
 ACAACCTCTA CTCAGTTCTA CTACAACTA TCTCAAGAAA TAAATGGTGA CATGGAACAG 1080  
 10 GTCAGTACT CCCTGGTCAC CTTGCAAGAT CAACTTAACT CCCTAGCAGC AGTAGTCCTT 1140  
 CAAAATCGAA GAGCTTTAGA CTTGCTAACC GCCAAAAGAG GGGGAACCTG TTTATTTTTA 1200  
 GGAGAAGAAC GCTGTTATTA TGTTAATCAA TCCAGAATTG TCACTGAGAA AGTTAAAGAA 1260  
 ATTCGAGATC GAATACAATG TAGAGCAGAG GAGCTTCAA ACACCGAACG CTGGGGCCTC 1320  
 CTCAGCCAAT GGATGCCCTG GGTCTCCCC TTCTTAGGAC CTCTAGCAGC TCTAATATTG 1380  
 15 TTACTCTCT TTGGACCCTG TATCTTTAAC CTCCTTGTTA AGTTTGTCTC TTCCAGAATT 1440  
 GAAGCTGTAA AGCTACAGAT GGTCTTACAA ATGGAACCCC A 1481

## (2) INFORMATION FOR SEQ ID NO: 118:

## (i) SEQUENCE CHARACTERISTICS:

- 20 (A) LENGTH: 493 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) TYPE DE MOLECULE: peptide

## 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 118:

Met Ala Leu Pro Tyr His Thr Phe Leu Phe Thr Val Leu Leu Pro Pro  
 1 5 10 15  
 Phe Ala Leu Thr Ala Pro Pro Pro Cys Cys Cys Thr Thr Ser Ser Ser  
 20 25 30  
 30 Pro Tyr Gln Glu Phe Leu Xaa Arg Thr Arg Leu Pro Gly Asn Ile Asp  
 35 40 45  
 Ala Pro Ser Tyr Arg Ser Leu Ser Lys Gly Asn Ser Thr Phe Thr Ala  
 50 55 60  
 His Thr His Met Pro Arg Asn Cys Tyr Asn Ser Ala Thr Leu Cys Met  
 35 65 70 75 80  
 His Ala Asn Thr His Tyr Trp Thr Gly Lys Met Ile Asn Pro Ser Cys

186

		85		90		95
		Pro Gly Gly Leu Gly Ala Thr Val Cys Trp Thr Tyr Phe Thr His Thr				
		100		105		110
		Ser Met Ser Asp Gly Gly Gly Ile Gln Gly Gln Ala Arg Glu Lys Gln				
5		115		120		125
		Val Lys Glu Ala Ile Ser Gln Leu Thr Arg Gly His Ser Thr Pro Ser				
		130		135		140
		Pro Tyr Lys Gly Leu Val Leu Ser Lys Leu His Glu Thr Leu Arg Thr				
		145		150		155
10		His Thr Arg Leu Val Ser Leu Phe Asn Thr Thr Leu Thr Arg Leu His				
		165		170		175
		Glu Val Ser Ala Gln Asn Pro Thr Asn Cys Trp Met Cys Leu Pro Leu				
		180		185		190
		His Phe Arg Pro Tyr Ile Ser Ile Pro Val Pro Glu Gln Trp Asn Asn				
15		195		200		205
		Phe Ser Thr Glu Ile Asn Thr Thr Ser Val Leu Val Gly Pro Leu Val				
		210		215		220
		Ser Asn Leu Glu Ile Thr His Thr Ser Asn Leu Thr Cys Val Lys Phe				
		225		230		235
20		Ser Asn Thr Ile Asp Thr Thr Ser Ser Gln Cys Ile Arg Trp Val Thr				
		245		250		255
		Pro Pro Thr Arg Ile Val Cys Leu Pro Ser Gly Ile Phe Phe Val Cys				
		260		265		270
		Gly Thr Ser Ala Tyr His Cys Leu Asn Gly Ser Ser Glu Ser Met Cys				
25		275		280		285
		Phe Leu Ser Phe Leu Val Pro Pro Met Thr Ile Tyr Thr Glu Gln Asp				
		290		295		300
		Leu Tyr Asn His Val Val Pro Lys Pro His Asn Lys Arg Val Pro Ile				
		305		310		315
30		Leu Pro Phe Val Ile Arg Ala Gly Val Leu Gly Arg Leu Gly Thr Gly				
		325		330		335
		Ile Gly Ser Ile Thr Thr Ser Thr Gln Phe Tyr Tyr Lys Leu Ser Gln				
		340		345		350
		Glu Ile Asn Gly Asp Met Glu Gln Val Thr Asp Ser Leu Val Thr Leu				
35		355		360		365
		Gln Asp Gln Leu Asn Ser Leu Ala Ala Val Val Leu Gln Asn Arg Arg				



187

370                      375                      380  
 Ala Leu Asp Leu Leu Thr Ala Lys Arg Gly Gly Thr Cys Leu Phe Leu  
 385                      390                      395                      400  
 Gly Glu Glu Arg Cys Tyr Tyr Val Asn Gln Ser Arg Ile Val Thr Glu  
 5                      405                      410                      415  
 Lys Val Lys Glu Ile Arg Asp Arg Ile Gln Cys Arg Ala Glu Glu Leu  
                     420                      425                      430  
 Gln Asn Thr Glu Arg Trp Gly Leu Leu Ser Gln Trp Met Pro Trp Val  
                     435                      440                      445  
 10 Leu Pro Phe Leu Gly Pro Leu Ala Ala Leu Ile Leu Leu Leu Leu Phe  
                     450                      455                      460  
 Gly Pro Cys Ile Phe Asn Leu Leu Val Lys Phe Val Ser Ser Arg Ile  
 465                      470                      475                      480  
 Glu Ala Val Lys Leu Gln Met Val Leu Gln Met Glu Pro  
 15                      485                      490

## (2) INFORMATION FOR SEQ ID NO: 119:

## (i) SEQUENCE CHARACTERISTICS:

- 20 (A) LENGTH: 32 base pairs  
 (B) TYPE: nucleotide  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) TYPE DE MOLECULE: ADNc

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 119:

25 TCAAAATCGA AGAGCTTTAG ACTTGCTAAC CG

32

## (2) INFORMATION FOR SEQ ID NO: 120:

## (i) SEQUENCE CHARACTERISTICS:

- 30 (A) LENGTH: 1329 base pairs  
 (B) TYPE: nucleotide  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) TYPE DE MOLECULE: ADNc

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 120:

35 TCAAAATCGA AGAGCTTTAG ACTTGCTAAC CGCCAAAAGA GGGGGAACCT GTTTATTTTT 60  
 AGGGGAAGAA TGCTGTTAGT ATGTTAATCA ATCTGGAATC ATTACTGAGA AAGTTAAAGA 120

188

```

AATTTGAGAT CGAATATAAT GTAGAGCAGA GGACCTTCAA AACACTGCAC CCTGGGGCCT 180
CCTCAGCCAA TGGATGCCCT GGACTCTCCC CTTCTTAGGA CCTCTAGCAG CTATAATATT 240
TTTACTCCTC TTTGGACCCT GTATCTTCAA CTTCTTGTT AAGTTTGTCT CTTCCAGAAT 300
TGAAGCTGTA AAGCTACAAA TAGTTCTTCA AATGGAACCC CAGATGCAGT CCATGACTAA 360
5 AATCTACCGT GGACCCCTGG ACCGGCCTGC TAGACTATGC TCTGATGTTA ATGACATTGA 420
AGTCACCCCT CCCGAGGAAA TCTCAACTGC ACAACCCCTA CTACACTCCA ATTCAGTAGG 480
AAGCAGTTAG AGCAGTTGTC AGCCAACCTC CCCAACAGTA CTTGGGTTTT CTTGTTGAGA 540
GGGTGGACTG AGAGACAGGA CTAGCTGGAT TTCCTAGGCT GACTAAGAAT CCCNAAGCCT 600
ANCTGGGAAG GTGACCGCAT CCATCTTTAA ACATGGGGCT TGCAACTTAG CTCACACCCG 660
10 ACCAATCAGA GAGCTCACTA AAATGCTAAT CAGGCAAAAA CAGGAGGTAA AGCAATAGCC 720
AATCATCTAT TGCCTGAGAG CACAGCGGGA AGGACAAGGA TTGGGATATA AACTCAGGCA 780
TTCAAGCCAG CAACAGCAAC CCCCTTTGGG TCCCCTCCCA TTGTATGGGA GCTCTGTTTT 840
CACTCTATTT CACTCTATTA AATCATGCAA CTGCACTCTT CTGGTCCGTG TTTTTTATGG 900
CTCAAGCTGA GCTTTTGTTC GCCATCCACC ACTGCTGTTT GCCACCGTCA CAGACCCGCT 960
15 GCTGACTTCC ATCCCTTTGG ATCCAGCAGA GTGTCCACTG TGCTCCTGAT CCAGCGAGGT 1020
ACCCATTGCC ACTCCCGATC AGGCTAAAGG CTTGCCATTG TTCCTGCATG GCTAAGTGCC 1080
TGGGTTTGTC CTAATAGAAC TGAACACTGG TCACTGGGTT CCATGGTTCT CTTCCATGAC 1140
CCACGGCTTC TAATAGAGCT ATAACACTCA CCGCATGGCC CAAGATTCCA TTCCTTGTA 1200
TCTGTGAGGC CAAGAACCCC AGGTCAGAGA ANGTGAGGCT TGCCACCATT TGGGAAGTGG 1260
20 CCCACTGCCA TTTTGGTAGC GGCCCACCAC CATCTTGGGA GCTGTGGGAG CAAGGATCCC 1320
CCAGTAACA 1329

```

## (2) INFORMATION FOR SEQ ID NO: 121:

## (i) SEQUENCE CHARACTERISTICS:

- 25 (A) LENGTH: 162 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) TYPE DE MOLECULE: peptide

## 30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 121:

```

Gln Asn Arg Arg Ala Leu Asp Leu Leu Thr Ala Lys Arg Gly Gly Thr
1           5           10           15
Cys Leu Phe Leu Gly Glu Glu Cys Cys Xaa Tyr Val Asn Gln Ser Gly
20           25           30
35 Ile Ile Thr Glu Lys Val Lys Glu Ile Xaa Asp Arg Ile Xaa Cys Arg
35           40           45

```

189

Ala Glu Asp Leu Gln Asn Thr Ala Pro Trp Gly Leu Leu Ser Gln Trp  
 50 55 60  
 Met Pro Trp Thr Leu Pro Phe Leu Gly Pro Leu Ala Ala Ile Ile Phe  
 65 70 75 80  
 5 Leu Leu Leu Phe Gly Pro Cys Ile Phe Asn Phe Leu Val Lys Phe Val  
 85 90 95  
 Ser Ser Arg Ile Glu Ala Val Lys Leu Gln Ile Val Leu Gln Met Glu  
 100 105 110  
 Pro Gln Met Gln Ser Met Thr Lys Ile Tyr Arg Gly Pro Leu Asp Arg  
 10 115 120 125  
 Pro Ala Arg Leu Cys Ser Asp Val Asn Asp Ile Glu Val Thr Pro Pro  
 130 135 140  
 Glu Glu Ile Ser Thr Ala Gln Pro Leu Leu His Ser Asn Ser Val Gly  
 145 150 155 160  
 15 Ser Ser

## (2) INFORMATION FOR SEQ ID NO: 122:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs  
 20 (B) TYPE: nucleotide  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) TYPE DE MOLECULE: ADNc

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 122:

25 GGCATTGATA GCACCCATCA G

21

## (2) INFORMATION FOR SEQ ID NO: 123:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs  
 30 (B) TYPE: nucleotide  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) TYPE DE MOLECULE: ADNc

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 123:

35 CATGTCACCA GGGTGAATA G

21

190

## (2) INFORMATION FOR SEQ ID NO: 124:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 758 base pairs

(B) TYPE: nucleotide

5 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) TYPE DE MOLECULE: ADNc

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 124:

GGCATTGATA GCACCCATCA GATGGCCAAA TCATTATTTA CTGGACCAGG CCTTTTCAAA 60  
10 ACTATCAAGC AGATAGGGCC CGTGAAGCAT GCCAAAGAAA TAATCCCCTG CCTTATCGCC 120  
ATGTTCTTC AGGAGAACAA AGAACAGGCC ATTACCCAGG GGAAGACTGG CAACTAGATT 180  
TTACCCACAT GGCCAAATGT CAGGGATTTC AGCATCTACT AGTCTGGGCA GATACTTTCA 240  
CTGGTTGGGT GGAGTCTTCT CCTTG TAGGA CAGAAAAGAC CCAAGAGGTA ATAAAGGCAC 300  
TAATGAAATA ATTCCCAGAT TTGGACTTCC CCCAGGATTA CAGGGTGACA ATGGCCCCGC 360  
15 TTTCAAGGCT GCAGTAACCC AGGGAGTATC CCAGGTGTTA GGCATACAAT ATCACTTACA 420  
CTGTGCCTGG AGGCCACAAT CCTCCAGAAA AGTCAAGAAA ATGAATGAAA CACTCAAAGA 480  
TCTAAAAAAG CTAACCCAAG AAACCCACAT TGCATGACCT GTTCTGTTGC CTATAACCTT 540  
ACTAAGAATC CATACTATC CCCCAAAAAG CAGGACTTAG CCCATACGAG ATGCTATATG 600  
GATGGCCTTT CCTAACCAAT GACCTTGTC TTGACTGAGA AATGGCCAAC TTAGTTGCAG 660  
20 ACATCACCTC CTTAGCCAAA TATCAACAAG TTCTTAAAC ATCACAGGGA ACCTGTCCCC 720  
GAGAGGAGGG AAAGGAACTA TTCCACCCTG GTGACATG 758

## (2) INFORMATION FOR SEQ ID NO: 126:

## 25 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

30 (ii) TYPE DE MOLECULE: ADNc

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 126:

CGGACATCCA AAGTGATGGG AAACG

25

## (2) INFORMATION FOR SEQ ID NO: 127:

## 35 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 base pairs

191

(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear  
(ii) TYPE DE MOLECULE: ADNc  
5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 127:  
GGACAGGAAA GTAAGACTGA GAAGGC 26

(2) INFORMATION FOR SEQ ID NO: 128:  
(i) SEQUENCE CHARACTERISTICS:  
10 (A) LENGTH: 26 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear  
(ii) TYPE DE MOLECULE: ADNc  
15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 128:  
CCTAGAACGT ATTCTGGAGA ATTGGG 26

(2) INFORMATION FOR SEQ ID NO: 129:  
(i) SEQUENCE CHARACTERISTICS:  
20 (A) LENGTH: 26 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear  
(ii) TYPE DE MOLECULE: ADNc  
25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129:  
TGGCTCTCAA TGGTCAAACA TACCCG 26

(2) INFORMATION FOR SEQ ID NO: 130:  
(i) SEQUENCE CHARACTERISTICS:  
30 (A) LENGTH: 1511 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear  
(ii) TYPE DE MOLECULE: ADNc  
35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 130:  
CCTAGAACGT ATTCTGGAGA ATTGGGACCA ATGTGACACT CAGACGCTAA GAAAGAAACG 60

192

```

      ATTTATATTC TTCTGCAGTA CCGCCTGGCC ACAATATCCT CTTCAAGGGA GAGAAACCTG 120
      GCTTCCTGAG GGAAGTATAA ATTATAACAT CATCTTACAG CTAGACCTCT TCTGTAGAAA 180
      GGAGGGCAAA TGGAGTGAAG TGCCATATGT GCAAACCTTC TTTTCATTAA GAGACAACTC 240
      ACAATTATGT AAAAAGTGTG GTTTATGCCC TACAGGAAGC CCTCAGAGTC CACCTCCCTA 300
5     CCCCAGCGTC CCCTCCCCGA CTCCTTCCTC AACTAATAAG GACCCCCCTT TAACCCAAAC 360
      GGTCCAAAAG GAGATAGACA AAGGGGTAAA CAATGAACCA AAGAGTGCCA ATATTCCCCG 420
      ATTATGCCCC CTCCAAGCAG TGAGAGGAGG AGAATTCGGC CCAGCCAGAG TGCCTGTACC 480
      TTTTCTCTC TCAGACTTAA AGCAAATTAA AATAGACCTA GGTAAATTCT CAGATAACCC 540
      TGACGGCTAT ATTGATGTTT TACAAGGGTT AGGACAATCC TTTGATCTGA CATGGAGAGA 600
10    TATAATGTTA CTACTAAATC AGACACTAAC CCCAAATGAG AGAAGTGCCG CTGTAAGTGC 660
      AGCCCGAGAG TTTGGCGATC TTTGGTATCT CAGTCAGGCC AACAATAGGA TGACAACAGA 720
      GGAAAGAACA ACTCCACAG GCCAGCAGGC AGTTCCCAGT GTAGACCCTC ATTGGGACAC 780
      AGAATCAGAA CATGGAGATT GGTGCCACAA ACATTGCTA ACTTGCGTGC TAGAAGGACT 840
      GAGGAAAAC AGGAAGAAGC CTATGAATTA CTCAATGATG TCCACTATAA CACAGGGAAA 900
15    GGAAGAAAAT CTTACTGCTT TTCTGGACAG ACTAAGGGAG GCATTGAGGA AGCATACCTC 960
      CCTGTCACCT GACTCTATTG AAGGCCAACT AATCTTAAAG GATAAGTTTA TCACTCAGTC 1020
      AGCTGCAGAC ATTAGAAAAA ACTTCAAAG TCTGCCTTAG GCCCGGAGCA GAACTTAGAA 1080
      ACCCTATTTA ACTTGGCATC CTCAGTTTTT TATAATAGAG ATCAGGAGGA GCAGGCGAAA 1140
      CGGGACAAAC GGGATAAAAA AAAAAGGGGG GGTCCACTAC TTTAGTCATG GCCCTCAGGC 1200
20    AAGCAGACTT TGGAGGCTCT GCAAAGGGA AAAGCTGGGC AAATCAAATG CCTAATAGGG 1260
      CTGGCTTCCA GTGCGGTCTA CAAGGACACT TTAAAAAGA TTATCCAAGT AGAAATAAGC 1320
      CGCCCCCTTG TCCATGCCCC TTACGTCAAG GGAATCACTG GAAGGCCAC TGCCCCAGGG 1380
      GATGAAGATA CTCTGAGTCA GAAGCCATTA ACCAGATGAT CCAGCAGCAG GACTGAGGGT 1440
      GCCCCGGGCG AGCGCCAGCC CATGCCATCA CCCTCACAGA GCCCCGGGTA TGTTTGACCA 1500
25    TTGAGAGCCA A                                         1511

```

(2) INFORMATION FOR SEQ ID NO: 131:

(i) SEQUENCE CHARACTERISTICS:

- 30 (A) LENGTH: 352 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) TYPE DE MOLECULE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 131:

35 Leu Glu Arg Ile Leu Glu Asn Trp Asp Gln Cys Asp Thr Gln Thr Leu  
 1 5 10 15

193

	Arg	Lys	Lys	Arg	Phe	Ile	Phe	Phe	Cys	Ser	Thr	Ala	Trp	Pro	Gln	Tyr
				20					25					30		
	Pro	Leu	Gln	Gly	Arg	Glu	Thr	Trp	Leu	Pro	Glu	Gly	Ser	Ile	Asn	Tyr
			35					40					45			
5	Asn	Ile	Ile	Leu	Gln	Leu	Asp	Leu	Phe	Cys	Arg	Lys	Glu	Gly	Lys	Trp
		50					55					60				
	Ser	Glu	Val	Pro	Tyr	Val	Gln	Thr	Phe	Phe	Ser	Leu	Arg	Asp	Asn	Ser
	65					70				75					80	
	Gln	Leu	Cys	Lys	Lys	Cys	Gly	Leu	Cys	Pro	Thr	Gly	Ser	Pro	Gln	Ser
10					85					90					95	
	Pro	Pro	Pro	Tyr	Pro	Ser	Val	Pro	Ser	Pro	Thr	Pro	Ser	Ser	Thr	Asn
				100					105					110		
	Lys	Asp	Pro	Pro	Leu	Thr	Gln	Thr	Val	Gln	Lys	Glu	Ile	Asp	Lys	Gly
			115				120						125			
15	Val	Asn	Asn	Glu	Pro	Lys	Ser	Ala	Asn	Ile	Pro	Arg	Leu	Cys	Pro	Leu
		130					135					140				
	Gln	Ala	Val	Arg	Gly	Gly	Glu	Phe	Gly	Pro	Ala	Arg	Val	Pro	Val	Pro
	145					150				155					160	
	Phe	Ser	Leu	Ser	Asp	Leu	Lys	Gln	Ile	Lys	Ile	Asp	Leu	Gly	Lys	Phe
20					165					170					175	
	Ser	Asp	Asn	Pro	Asp	Gly	Tyr	Ile	Asp	Val	Leu	Gln	Gly	Leu	Gly	Gln
			180						185					190		
	Ser	Phe	Asp	Leu	Thr	Trp	Arg	Asp	Ile	Met	Leu	Leu	Leu	Asn	Gln	Thr
			195				200							205		
25	Leu	Thr	Pro	Asn	Glu	Arg	Ser	Ala	Ala	Val	Thr	Ala	Ala	Arg	Glu	Phe
		210					215					220				
	Gly	Asp	Leu	Trp	Tyr	Leu	Ser	Gln	Ala	Asn	Asn	Arg	Met	Thr	Thr	Glu
	225					230				235					240	
	Glu	Arg	Thr	Thr	Pro	Thr	Gly	Gln	Gln	Ala	Val	Pro	Ser	Val	Asp	Pro
30					245					250					255	
	His	Trp	Asp	Thr	Glu	Ser	Glu	His	Gly	Asp	Trp	Cys	His	Lys	His	Leu
			260						265					270		
	Leu	Thr	Cys	Val	Leu	Glu	Gly	Leu	Arg	Lys	Thr	Arg	Lys	Lys	Pro	Met
			275					280					285			
35	Asn	Tyr	Ser	Met	Met	Ser	Thr	Ile	Thr	Gln	Gly	Lys	Glu	Glu	Asn	Leu
		290							295					300		

194

Thr Ala Phe Leu Asp Arg Leu Arg Glu Ala Leu Arg Lys His Thr Ser  
 305 310 315 320  
 Leu Ser Pro Asp Ser Ile Glu Gly Gln Leu Ile Leu Lys Asp Lys Phe  
 325 330 335  
 5 Ile Thr Gln Ser Ala Ala Asp Ile Arg Lys Asn Phe Lys Ser Leu Pro  
 340 345 350

## (2) INFORMATION FOR SEQ ID NO: 132:

## (i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 30 base pairs  
 (B) TYPE: nucleotide  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) TYPE DE MOLECULE: ADNc

- 15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 132:

TGCTGGAATT CGGGATCCTA GAACGTATTC

30

## (2) INFORMATION FOR SEQ ID NO: 133:

## (i) SEQUENCE CHARACTERISTICS:

- 20 (A) LENGTH: 30 base pairs  
 (B) TYPE: nucleotide  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) TYPE DE MOLECULE: ADNc

- 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 133:

AGTTCTGCTC CGAAGCTTAG GCAGACTTTT

30

## (2) INFORMATION FOR SEQ ID NO: 135:

## (i) SEQUENCE CHARACTERISTICS:

- 30 (A) LENGTH: 398 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) TYPE DE MOLECULE: peptide

- 35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 135:

Met Gly Ser Ser His His His His His His Ser Ser Gly Leu Val Pro



195

	1		5		10		15									
	Arg	Gly	Ser	His	Met	Ala	Ser	Met	Thr	Gly	Gly	Gln	Gln	Met	Gly	Arg
				20				25						30		
	Ile	Leu	Glu	Arg	Ile	Leu	Glu	Asn	Trp	Asp	Gln	Cys	Asp	Thr	Gln	Thr
5			35				40				45					
	Leu	Arg	Lys	Lys	Arg	Phe	Ile	Phe	Phe	Cys	Ser	Thr	Ala	Trp	Pro	Gln
		50				55				60						
	Tyr	Pro	Leu	Gln	Gly	Arg	Glu	Thr	Trp	Leu	Pro	Glu	Gly	Ser	Ile	Asn
	65				70					75					80	
10	Tyr	Asn	Ile	Ile	Leu	Gln	Leu	Asp	Leu	Phe	Cys	Arg	Lys	Glu	Gly	Lys
				85					90					95		
	Trp	Ser	Glu	Val	Pro	Tyr	Val	Gln	Thr	Phe	Phe	Ser	Leu	Arg	Asp	Asn
			100					105					110			
	Ser	Gln	Leu	Cys	Lys	Lys	Cys	Gly	Leu	Cys	Pro	Thr	Gly	Ser	Pro	Gln
15			115				120					125				
	Ser	Pro	Pro	Pro	Tyr	Pro	Ser	Val	Pro	Ser	Pro	Thr	Pro	Ser	Ser	Thr
		130				135					140					
	Asn	Lys	Asp	Pro	Pro	Leu	Thr	Gln	Thr	Val	Gln	Lys	Glu	Ile	Asp	Lys
	145				150					155					160	
20	Gly	Val	Asn	Asn	Glu	Pro	Lys	Ser	Ala	Asn	Ile	Pro	Arg	Leu	Cys	Pro
			165						170				175			
	Leu	Gln	Ala	Val	Arg	Gly	Gly	Glu	Phe	Gly	Pro	Ala	Arg	Val	Pro	Val
			180					185				190				
	Pro	Phe	Ser	Leu	Ser	Asp	Leu	Lys	Gln	Ile	Lys	Ile	Asp	Leu	Gly	Lys
25			195				200					205				
	Phe	Ser	Asp	Asn	Pro	Asp	Gly	Tyr	Ile	Asp	Val	Leu	Gln	Gly	Leu	Gly
		210				215					220					
	Gln	Ser	Phe	Asp	Leu	Thr	Trp	Arg	Asp	Ile	Met	Leu	Leu	Leu	Asn	Gln
	225				230				235					240		
30	Thr	Leu	Thr	Pro	Asn	Glu	Arg	Ser	Ala	Ala	Val	Thr	Ala	Ala	Arg	Glu
			245					250				255				
	Phe	Gly	Asp	Leu	Trp	Tyr	Leu	Ser	Gln	Ala	Asn	Asn	Arg	Met	Thr	Thr
		260				265				270						
	Glu	Glu	Arg	Thr	Thr	Pro	Thr	Gly	Gln	Gln	Ala	Val	Pro	Ser	Val	Asp
35			275				280					285				
	Pro	His	Trp	Asp	Thr	Glu	Ser	Glu	His	Gly	Asp	Trp	Cys	His	Lys	His

196

290                      295                      300  
 Leu Leu Thr Cys Val Leu Glu Gly Leu Arg Lys Thr Arg Lys Lys Pro  
 305                      310                      315                      320  
 Met Asn Tyr Ser Met Met Ser Thr Ile Thr Gln Gly Lys Glu Glu Asn  
 5                      325                      330                      335  
 Leu Thr Ala Phe Leu Asp Arg Leu Arg Glu Ala Leu Arg Lys His Thr  
 340                      345                      350  
 Ser Leu Ser Pro Asp Ser Ile Glu Gly Gln Leu Ile Leu Lys Asp Lys  
 355                      360                      365  
 10 Phe Ile Thr Gln Ser Ala Ala Asp Ile Arg Lys Asn Phe Lys Ser Leu  
 370                      375                      380  
 Pro Lys Leu Ala Ala Ala Leu Glu His His His His His His  
 385                      390                      395

## 15 (2) INFORMATION FOR SEQ ID NO: 137:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 378 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 20 (D) TOPOLOGY: linear

## (ii) TYPE DE MOLECULE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 137:

Met Ala Ser Met Thr Gly Gly Gln Gln Met Gly Arg Ile Leu Glu Arg  
 1                      5                      10                      15  
 25 Ile Leu Glu Asn Trp Asp Gln Cys Asp Thr Gln Thr Leu Arg Lys Lys  
 20                      25                      30  
 Arg Phe Ile Phe Phe Cys Ser Thr Ala Trp Pro Gln Tyr Pro Leu Gln  
 35                      40                      45  
 Gly Arg Glu Thr Trp Leu Pro Glu Gly Ser Ile Asn Tyr Asn Ile Ile  
 30                      50                      55                      60  
 Leu Gln Leu Asp Leu Phe Cys Arg Lys Glu Gly Lys Trp Ser Glu Val  
 65                      70                      75                      80  
 Pro Tyr Val Gln Thr Phe Phe Ser Leu Arg Asp Asn Ser Gln Leu Cys  
 85                      90                      95  
 35 Lys Lys Cys Gly Leu Cys Pro Thr Gly Ser Pro Gln Ser Pro Pro Pro  
 100                      105                      110

197

Tyr Pro Ser Val Pro Ser Pro Thr Pro Ser Ser Thr Asn Lys Asp Pro  
 115 120 125  
 Pro Leu Thr Gln Thr Val Gln Lys Glu Ile Asp Lys Gly Val Asn Asn  
 130 135 140  
 5 Glu Pro Lys Ser Ala Asn Ile Pro Arg Leu Cys Pro Leu Gln Ala Val  
 145 150 155 160  
 Arg Gly Gly Glu Phe Gly Pro Ala Arg Val Pro Val Pro Phe Ser Leu  
 165 170 175  
 Ser Asp Leu Lys Gln Ile Lys Ile Asp Leu Gly Lys Phe Ser Asp Asn  
 10 180 185 190  
 Pro Asp Gly Tyr Ile Asp Val Leu Gln Gly Leu Gly Gln Ser Phe Asp  
 195 200 205  
 Leu Thr Trp Arg Asp Ile Met Leu Leu Leu Asn Gln Thr Leu Thr Pro  
 210 215 220  
 15 Asn Glu Arg Ser Ala Ala Val Thr Ala Ala Arg Glu Phe Gly Asp Leu  
 225 230 235 240  
 Trp Tyr Leu Ser Gln Ala Asn Asn Arg Met Thr Thr Glu Glu Arg Thr  
 245 250 255  
 Thr Pro Thr Gly Gln Gln Ala Val Pro Ser Val Asp Pro His Trp Asp  
 20 260 265 270  
 Thr Glu Ser Glu His Gly Asp Trp Cys His Lys His Leu Leu Thr Cys  
 275 280 285  
 Val Leu Glu Gly Leu Arg Lys Thr Arg Lys Lys Pro Met Asn Tyr Ser  
 290 295 300  
 25 Met Met Ser Thr Ile Thr Gln Gly Lys Glu Glu Asn Leu Thr Ala Phe  
 305 310 315 320  
 Leu Asp Arg Leu Arg Glu Ala Leu Arg Lys His Thr Ser Leu Ser Pro  
 325 330 335  
 Asp Ser Ile Glu Gly Gln Leu Ile Leu Lys Asp Lys Phe Ile Thr Gln  
 30 340 345 350  
 Ser Ala Ala Asp Ile Arg Lys Asn Phe Lys Ser Leu Pro Lys Leu Ala  
 355 360 365  
 Ala Ala Leu Glu His His His His His His  
 370 375

35

(2) INFORMATION FOR SEQ ID NO: 138:

198

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 25 base pairs  
    (B) TYPE: nucleotide  
    (C) STRANDEDNESS: single  
5      (D) TOPOLOGY: linear  
    (ii) TYPE DE MOLECULE: ADNc  
    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 138:  
        CTTGGAGGGT GCATAACCAG GGAAT 25
- 10 (2) INFORMATION FOR SEQ ID NO: 139:  
    (i) SEQUENCE CHARACTERISTICS:  
        (A) LENGTH: 20 base pairs  
        (B) TYPE: nucleotide  
        (C) STRANDEDNESS: single  
15      (D) TOPOLOGY: linear  
    (ii) TYPE DE MOLECULE: ADNc  
    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 139:  
        TGTCCGCTGT GCTCCTGATC 20
- 20 (2) INFORMATION FOR SEQ ID NO: 140:  
    (i) SEQUENCE CHARACTERISTICS:  
        (A) LENGTH: 25 base pairs  
        (B) TYPE: nucleotide  
        (C) STRANDEDNESS: single  
25      (D) TOPOLOGY: linear  
    (ii) TYPE DE MOLECULE: ADNc  
    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 140:  
        CTATGTCCTT TTGGACTGTT TGGGT 25
- 30 (2) INFORMATION FOR SEQ ID NO: 141:  
    (i) SEQUENCE CHARACTERISTICS:  
        (A) LENGTH: 764 base pairs  
        (B) TYPE: nucleotide  
        (C) STRANDEDNESS: single  
35      (D) TOPOLOGY: linear  
    (ii) TYPE DE MOLECULE: ADNc

199

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 141:

TGTCCGCTGT GCTCCTGATC CAGCACAGGC GCCCATTGCC TCTCCCAATT GGGCTAAAGG 60  
 CTTGCCATTG TTCCTGCACA GCTAAGTGCC TGGGTTTCATC CTAATCGAGC TGAACACTAG 120  
 TCACTGGGTT CCACGGTTCT CTTCCATGAC CCATGGCTTC TAATAGAGCT ATAACACTCA 180  
 5 CTGCATGGTC CAAGATTCCA TTCCTTGGA TCCGTGAGAC CAAGAACCCC AGGTCAGAGA 240  
 ACACAAGGCT TGCCACCATG TTGGAAGCAG CCCACCACCA TTTTGGAAGC AGCCCGCCAC 300  
 TATCTTGGGA GCTCTGGGAG CAAGGACCCC AGGTAACAAT TTGGTGACCA CGAAGGGACC 360  
 TGAATCCGCA ACCATGAAGG GATCTCCAAA GCAATTGGAA ATGTTCTCTC CAAGGCAAAA 420  
 ATGCCCCTAA GATGTATTCT GGAGAATTGG GACCAATTTG ACCCTCAGAC AGTAAGAAAA 480  
 10 AAATGACTTA TATTCTTCTG CAGTACCGCC CTGGCCACGA TATCCTCTTC AAGGGGGAGA 540  
 AACCTGGCCT CCTGAGGGAA GTATAAATTA TAACACCATC TTACAGCTAG ACCTGTTTTG 600  
 TAGAAAAGGA GGCAAAATGGA GTGAAGTGCC ATATTTACAA ACTTTCTTTT CATTAAAAGA 660  
 CAACTCGCAA TTATGTTAAC AGTGTGATTT GTGTTCTTAC ACGGAAGCCC TCAGATTCTA 720  
 CTCCCCACCC CCGGCATCTC CCCTGAATCC CTCCCCAACT TATT 764  
 15

## (2) INFORMATION FOR SEQ ID NO: 142:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 800 base pairs  
 (B) TYPE: nucleotide  
 20 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) TYPE DE MOLECULE: ADNc

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 142:

TGTCCGCTGT GCTCCTGATC CAGCACAGGC GCCCATTGCC TCTCCCAATT GGGCTAAAGG 60  
 25 CTTGCCATTG TTCCTGCACA GCTAAGTGCC TGGGTTTCATC CTAATCGAGC TGAACACTAG 120  
 TCACTGGGTT CCACGGTTCT CTTCCATGAC CCATGGCTTC TAATAGAGCT ATAACACTCA 180  
 CTGCATGGTC CAAGATTCCA TTCCTTGGA TCCGTGAGAC CAAGAACCCC AGGTCAGAGA 240  
 ACACAAGGCT TGCCACCATG TTGGAAGCAG CCCACCACCA TTTTGGAAGC GGCCCGCCAC 300  
 TATCTTGGGA GCTCTGGGAG CAAGGACCCC CAGGTAACAA TTTGGTGACC ACGAAGGGAC 360  
 30 CTGAATCCGC AACCATGAAG GGATCTCCAA AGCAATTGGA AATGTTCTC CCAAGGCAAA 420  
 AATGCCCCTA AGATGTATTG TGGAGAATTG GGACCAATCT GACCCTCAGA CAGTAAGAAA 480  
 AAAAATGACT TATATTCTTC TGCAGTACCG CCTGGCCACG GATATCCTCT TCAAGGGGGA 540  
 GAAACCTGGC CTCCTGAGGG AAGTATAAAT TATAACACCA TCTTACAGCT AGACCTGTTT 600  
 TGTAAGAAAAG GAGGCAAATG GAGTGAAGTG CCATATTTAC AAACCTTTCTT TTCATTAAAA 660  
 35 GACAACTCGC AATTATGTAA ACAGTGTGAT TTGTGTCCTA CAGGAAGCCC TCAGATCTAC 720  
 CTCCCTACCC CGGCATCTCC CTGACTCCTT CCCCACCTAA TAAGGACCCA CTTAGCCCCA 780

200

AACAGTCCAA AAGGACATAG

800

## (2) INFORMATION FOR SEQ ID NO: 169:

## (i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (ii) TYPE DE MOLECULE: ADNc

## 10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:169:

consensus (41/68-1 + 42/68-1 + cl43 68-1)

## (2) INFORMATION FOR SEQ ID NO: 170:

## (i) SEQUENCE CHARACTERISTICS:

- 15 (A) LENGTH: 438 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (ii) TYPE DE MOLECULE: ADNc

## 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:170:

GACTTGAGCC AGTCCTCATA CCTGGACACT CTTGTCCTTC GGTACATGGA TGATTTACTT 60  
TTAGCCACCC ATTCAAGAAC CTTGTGCCAT CAAGCCACCC AAGCACTCTT AAATTTCCCTT 120  
GCTACCTGTG GCTACAAGGT TTCAAACCA AAGGCTCAGC TCTGCTCACA GCAGGTAAAA 180  
TACTTAGGGC TAAAATTATC CAAAGGCACC AGAACCTCA GTGAGGAACG TATCCAGCCT 240  
25 ATACTGGGTT ATCCTCATCC CAAAACCCTA AAGCAACTAA CAGCGTTCCT TGGCATAACA 300  
GGTTTCTGCC AAATATGGAT TCCCAGGTAC AGCAAGATAG CCAGACCATT AAATACACGA 360  
ATTAAGGAAA CTCAAAAAGC CAATACCCAT TTAGTAAGAT GGACACCTGA AGCAGAAGTG 420  
GCTTTCCAGG CCCTAAAG 438

## 30 (2) INFORMATION FOR SEQ ID NO: 171:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 438 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
35 (D) TOPOLOGY: linear

## (ii) TYPE DE MOLECULE: ADNc

201

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:171:

GACTTGAGCC AGTCCTCATA CCTGGACACT CTTGTCCTTC GGTACATGGA TGATTTACTT 60  
 TTAGCCACCC ATTCAGAAAC CTTGTGCCAT CAAGCCACCC AAGCACTCTT AAATTTTCCTT 120  
 GCTACCTGTG GCTACAAGGT TTCCAAACCA AAGGCTCAGC TCTGCTCACA GCAGGTAA 180  
 5 TACTTAGGGC TAAAATTATC CAAAGGCACC AGAACCTCA GTGAGGAACG TATCCAGCCT 240  
 ATACTGGGTT ATCCTCATCC CAAAACCCTA AAGCAACTAA CAGCGTTCCT TGGCATAACA 300  
 GGTTTCTGCC AAATATGGAT TCCCAGGTAC AGCAAAGTAG CCAGACCATT AAATACACGA 360  
 ATTAAGGAAA CTCAAAAAGC CAGTACCCAT TTAGTAAGAT GGACACCTGA AGCAGAAGTG 400  
 GCTTTCCAGG CCCTAAAG 438

10

## (2) INFORMATION FOR SEQ ID NO: 172:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 438 base pairs

(B) TYPE: nucleotide

15

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

## (ii) TYPE DE MOLECULE: ADNc

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:172:

GACTTGAGCC AGTCYTCATA CCTGGACAYT CTTGTCCTTC GGTACATGGA TGATTTACTT 60  
 20 TTAGCCACCC ATTCAGAAAC CTTGTGCCAT CAAGCCACCC AAGCACTCTT AAATTTTCCTT 120  
 GCTACCTGTG GCTACAAGGT TTCCAAACCA AAGGCTCAGC TCTGCTCACA GCAGGTAA 180  
 TACTTAGGGC TAAAATTATC CAAAGGCACC AGAACCTCA GTGAGGAACG TATCCAGCCT 240  
 ATACTGGGTT ATCCTCATCC CAAAACCCTA AAGCAACTAA CAGCGTTCCT TGGCATAACA 300  
 GGTTTCTGCC AAATATGGAT TCCCAGGTAC AGCAAAGTAG CCAGACCATT AAATACACGA 360  
 25 ATTAAGGAAA CTCAAAAAGC CAATACCCAT TTAGTAAGAT GGACATCTGA AGCAGAAGTG 400  
 GCTTTCCAGG CCCTAAAG 438

## (2) INFORMATION FOR SEQ ID NO: 173:

## (i) SEQUENCE CHARACTERISTICS:

30

(A) LENGTH: 146 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

## (ii) TYPE DE MOLECULE: peptide

35

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:173:

DLSQSSYLDL LVLRYMDDL LATHSETLCH QATQALLNFL ATCGYKVKSP 50

202

KAQLCSQQVK YLGLKLSKGT RTLSEERIQP ILGYPHPKTL KQLTAFLGIT 100  
GFCQIWIPRY SKIARPLNTR IKETQKANTH LVRWTPEAEV AFQALK 146

(2) INFORMATION FOR SEQ ID NO: 174:

5 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 146 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

10 (ii) TYPE DE MOLECULE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:174:

DLSQSSYLDL LVLRYMDDL LATHSETLCH QATQALLNFL ATCGYKVSKEP 50  
KAQLCSQQVK YLGLKLSKGT RTLSEERIQP ILGYPHPKTL KQLTAFLGIT 100  
GFCQIWIPRY SKVARPLNTR IKETQKASTH LVRWTPEAEV AFQALK 146

15

(2) INFORMATION FOR SEQ ID NO: 175:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 146 amino acids

(B) TYPE: amino acid

20 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) TYPE DE MOLECULE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:175:

DLSQSSYLDX LVLRYMDDL LATHSETLCH QATQALLNFL ATCGYKVSKEP 50  
25 KAQLCSQQVK YLGLKLSKGT RTLSEERIQP ILGYPHPKTL KQLTAFLGIT 100  
GFCQIWIPRY SKIARPLNTR IKETQKANTH LVRWTSEAEV AFQALK 146

(2) INFORMATION FOR SEQ ID NO: 176:

(i) SEQUENCE CHARACTERISTICS:

30 (A) LENGTH: base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) TYPE DE MOLECULE: ADNc

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:176:

consensus (1/46-7+8/46-7+c15/46/7)



203

## (2) INFORMATION FOR SEQ ID NO: 177:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 429 base pairs  
 5 (B) TYPE: nucleotide  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) TYPE DE MOLECULE: ADNc

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:177:

```

10 GACTTGAGCC AGTCCTCATA CCTGGACATT CTTGTTCTTC AGTATGGGGA TGA CTTAATT 60
   ATAGCCACCC ATTCAGAAAC CTTGTGGCAT CAAGCCACCC AAGCGCTCTT AAATTTCTT 120
   GCTACCTGTG GCTCCAAACA AAAGGCTCAC CTCTGCTCAC ACCAGGTAA ATACTTAGGG 180
   CTA AATTAT CCAAAGTCAC CAGGGCCCTC AGAGAGGAAC GTATCCAGCG TATACTGGCT 240
   TATCCTCATC CCATAACCCT AAAGCAACTA AGAGGGTTCC TTGGCATATC AGCCTTCTGC 300
15 CGAATATGGA TTCCCGGATA CAGTGAAATA GCCAGGCCAT TATGTACATT AATTAAGGAA 360
   ACTCAGAAAG CCAATACCCA TATAGTAAGA TGGACACCTG AAACAGAAGT GGCTTTCCAG 420
   GCCCTAAAG 429

```

## (2) INFORMATION FOR SEQ ID NO: 178:

## 20 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 429 base pairs  
 (B) TYPE: nucleotide  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## 25 (ii) TYPE DE MOLECULE: ADNc

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:178:

```

   GACTTGAGCC AGTCCTCATA CCTGGACATT CTTGTTCTTC AGTATAGGGA TGATTTAATT 60
   ATAGCCACCC ATTCAGAAAC CTTGTGGCAT CAAGCCACCC AAGTGCTCTT AAATTTCTTC 120
   GCTACCTGTG GCTCCAAACA AAGGGCTCAG CTCTGCTCAC AGCAGGTAA ATACTTAGGG 180
30 CTA AATTAT CCAAAGTCGC CAGGGCCCTC AGAGAGGAAC GTATCCAGCG TATACTGGAT 240
   TATCCTCATC CCAAACCAT AAAGCAACTA AGAGGGTTCC TTGGCATAAC AGCCTTCTGC 300
   CGAATATGGA TTCCCGGATA CAGTGAAATA GCCAGGCCAT TATGTACATT AGTTAAGGAA 360
   ACTCAGAAAG CCAATACCCA TATAGTAAGA TGGACACCTG AGACAGAAGT GGCTTTCCAG 420
   GCCCTAAAG 429

```

35

## (2) INFORMATION FOR SEQ ID NO: 179:

204

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 429 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

5 (D) TOPOLOGY: linear

## (ii) TYPE DE MOLECULE: ADNc

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:179:

GACTTGAGCC AGTCCTCATA CCTGGACATT CTTGTTCCTC AGTATGGGGA TGATTTAATT 60  
 ATAGCCACCC ATTCAAGAAC CTTGTGGCAC CAAGCCACCC AAGCGCTCTT AAATTTCTCTC 120  
 10 GCTACCTGTG GCTCCAAACA AAAGGCTCAG CTCTGCTCAC AGCAGGTAA ATACTTAGGG 180  
 CTAAAATTAT CCAAAGTCAC CAGGGCCCTC AGAGAGGAAC GTATCCAGCG TATACTGGCT 240  
 TATCCCCATC CCAAACCT AAAGCAACTA AGARGGTTCC TTGGCATAAC AGCCTTCTGC 300  
 CGAATATGGA TTCCAGATA CAGCGAATA GCCAGGCCAT TATGTACATT ATCTAAGGAA 360  
 ACTCAGAAAG CCAATACCCA TATAGTAAGA TGGACACCTG AACAGAAGT GGCTTTCCAG 420  
 15 GCCCTAAAG 429

## (2) INFORMATION FOR SEQ ID NO: 180:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 143 amino acids

20 (B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

## (ii) TYPE DE MOLECULE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:180:

25 DLSQSSYLDI LVLQYGDDLI IATHSETLWH QATQALLNFL ATCGSKQKAH 50  
 LCSHQVKYLG LKLSKVTRAL REERIQRILA YPHPITLKQL RGFLGISAFQ 100  
 RIWIPGYSEI ARPLCTLIKE TQKANTHIVR WTPETEVAFQ ALK 143

## (2) INFORMATION FOR SEQ ID NO: 181:

## 30 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 143 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

35 (ii) TYPE DE MOLECULE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:181:

205

DLSQSSYLDI LVLQYRDDLI IATHSETLWH QATQVLLNFL ATCGSKQRAQ 50  
LCSQQVKYLG LKLSKVARAL REERIQRILD YPHPKTIKQL RGFLGITAF 100  
RIWIPRYSEI ARPLCTLVKE TQKANTHIVR WTPETEVAFQ ALK 143

## 5 (2) INFORMATION FOR SEQ ID NO: 182:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 143 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

10 (D) TOPOLOGY: linear

(ii) TYPE DE MOLECULE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:182:

DLSQSSYLDI LVPQYGGDLI IATHSETLWH QATQALLNFL ATCGSKQKAQ 50  
LCSQQVKYLG LKLSKVTRAL REERIQRILA YPHPKTLKQL RXFLGITAF 100  
15 RIWIPRYSEI ARPLCTLKE TQKANTHIVR WTPETEVAFQ ALK 143

## (2) INFORMATION FOR SEQ ID NO: 183:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs

20 (B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) TYPE DE MOLECULE: ADNC

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:183:

25 GGCCAGGCAT CAGCCAAGA CTTGA 25

## (2) INFORMATION FOR SEQ ID NO: 184:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs

30 (B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) TYPE DE MOLECULE: ADNC

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:184:

35 TGCAAGCTCA TCCCTSRGAC CT 22

206

## (2) INFORMATION FOR SEQ ID NO: 185:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 base pairs

(B) TYPE: nucleotide

5 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) TYPE DE MOLECULE: ADNc

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:185:

GACTTGAGCC AGTCCTCATA CCT

23

10

## (2) INFORMATION FOR SEQ ID NO: 186:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs

(B) TYPE: nucleotide

15 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) TYPE DE MOLECULE: ADNc

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:186:

CTTTAGGGCC TGGAAAGCCA CT

22

20

TABLE No. 5

SEQUENCES GENERATED BY PAN-RETROVIRUS PCR OF DENSITY GRADIENT FRACTIONS  
(containing the peak of RT-activity or the corresponding control fraction)

CULTURE	MSRV c-pol	ERV8(v)	PCR artifacts (vI)	Total clones
LM7P (II)	18	4	8	28
PLI-1 (III)	0	1	13	23
MS B-CELL LINE (III)	9	2	8	19
CONTROL B-CELL LINE (IV)	0	0	28	28

- I LM7-infected choroid plexus cell culture .
- II MS patient-derived choroid plexus cell culture (PLI-2).
- III MS patient-derived spontaneous B-cell line (immortalized by endogenous EBV strain).
- IV Non-MS control B-cell line.
- V Clones with >90% homology with the GenBank sequence HSERV9 are designated ERV9 in this study.
- VI PCR artifacts included primer multimers, concatamers, single primer amplifications, etc.

TABLE No. 6

## DETECTION OF HMRV IN THE CSF OF PATIENTS WITH MULTIPLE SCLEROSIS AND OTHER NEUROLOGICAL DISEASES

Patient <sup>1</sup>	Age/Sex	Diagnosis	MS Type	MS Activity	MS Duration	MS Treatment at sampling	MSRV ELOSA
ITMS1	27 yrs / M	multiple sclerosis	2° progressive	slow progression	5 yrs	corticosteroids	negative
ITMS2	55 yrs / M	multiple sclerosis	1° progressive	slow progression	8 yrs	none	POSITIVE
ITMS3	51 yrs / F	multiple sclerosis	1° progressive	slow progression	2 yrs	none	negative
ITMS4	22 yrs / F	multiple sclerosis	relapsing / remitting	progression	0 yrs	none	POSITIVE
ITMS5	27 yrs / F	multiple sclerosis	1° progressive	slow progression	8 yrs	cyclophosphamide	negative
ITMS6	33 yrs / M	multiple sclerosis	2° progressive	slow progression	18 yrs	none (previously cycloph. + corticost.)	negative
ITMS7	33 yrs / F	multiple sclerosis	2° progressive	slow progression	8 yrs	none	POSITIVE
ITMS8	25 yrs / F	multiple sclerosis	relapsing / remitting	stable	3 yrs	none	POSITIVE
ITMS9	38 yrs / F	multiple sclerosis	2° progressive	slow progression	3 yrs	none	POSITIVE
ITMS10	38 yrs / M	multiple sclerosis	2° progressive	slow progression	7 yrs	corticosteroids	negative
OND1	37 yrs / F	cerebellar atrophy	NA <sup>2</sup>	NA	NA	NA	negative
OND2	26 yrs / F	viral myelitis	NA	NA	NA	NA	negative
OND3	38 yrs / F	viral encephalitis	NA	NA	NA	NA	negative
OND4	28 yrs / F	viral encephalitis	NA	NA	NA	NA	negative
OND5	54 yrs / M	viral encephalitis	NA	NA	NA	NA	negative
OND6	32 yrs / M	Gullain - Barré	NA	NA	NA	NA	negative
OND7	54 yrs / F	cerebrovascular	NA	NA	NA	NA	negative
OND8	52 yrs / F	hydrocephalus	NA	NA	NA	NA	negative
OND9	25 yrs / F	1° cerebral tumour	NA	NA	NA	NA	negative
OND10	21 yrs / M	epilepsy	NA	NA	NA	NA	negative

<sup>1</sup> CSF samples from patients ITMS1 - OND2 were made available by Prof. P. Ferrante, University Centre for Multiple Sclerosis, Milan, Italy.

CSF samples from patients OND3 - OND10 were made available by Profs. J. Pallat and J. Porrel, Dept. of Neurology, University Hospital, Grenoble, France.

<sup>2</sup> NA = Not Applicable

## CLAIMS

1. Nucleic material, in the isolated or purified state, comprising a nucleotide sequence selected  
5 from the group including sequences SEQ ID NO:93, SEQ ID NO:94, their complementary sequences and their equivalent sequences, in particular nucleotide sequences displaying, for any succession of 100 contiguous monomers, at least 50% and preferably at least 60% homology with said  
10 sequence SEQ ID NO:93, SEQ ID NO:94 and their complementary sequences, excluding HSERV-9 sequence.

2. Nucleic material of claim 1, nucleotide sequence of which is selected from the group including sequences SEQ ID NO:93, SEQ ID NO:94, their complementary  
15 sequences and their equivalent sequences, in particular nucleotide sequences displaying, for any succession of 100 contiguous monomers, at least 70% and preferably at least 80% homology with said sequence SEQ ID NO:93, SEQ ID NO:94 and their complementary sequences.

20 3. Nucleic material, in the isolated or purified state, coding for any polypeptide displaying, for any contiguous succession of at least 30 amino acids, at least 50%, preferably at least 60 %, and most preferably at least 70% homology with a peptide sequence encoded by  
25 any nucleotide sequence selected from the group including SEQ ID NO:93, SEQ ID NO:94 and their complementary sequence.

4. Nucleic material, in the isolated or purified state, of retroviral type, comprising a  
30 nucleotide sequence identical or equivalent to at least part of the pol gene of an isolated retrovirus associated with multiple sclerosis or rheumatoid arthritis.

5. Nucleic material as claimed in claim 4, said nucleotide sequence being 80 % homologous to said at  
35 least part of the pol gene.

6. Nucleic material comprising a nucleotide sequence identical or equivalent to at least part of the pol gene of an isolated virus encoding a reverse transcriptase comprising an enzymatic site comprised  
5 between the amino acid domains LPQG and YXDD, said virus having a phylogenic distance with HSERV-9 of  $0.063 \pm 0.1$ , and preferably  $0.063 \pm 0.05$ .

7. Nucleotide fragment comprising a nucleotide sequence selected from the group including SEQ ID NO:93,  
10 SEQ ID NO: 94, their complementary sequences and their equivalent sequences, in particular nucleotide sequences displaying, for any succession of 100 contiguous monomers, at least 50% and preferably at least 60% homology with said sequences and their complementary sequences, said  
15 group excluding SEQ ID NO:1, and said nucleotide fragment not comprising nor consisting of the sequence HSERV-9.

8. Nucleotide fragment of claim 7, nucleotide sequence of which is selected from the group including SEQ ID NO:93, SEQ ID NO: 94, their complementary sequences and  
20 their equivalent sequences, in particular nucleotide sequences displaying, for any succession of 100 contiguous monomers, at least 70% and preferably at least 80% homology with said sequences and their complementary sequences.

25 9. Nucleotide fragment comprising a coding nucleotide sequence which is at least partially identical to a nucleotide sequence selected from the group including :

SEQ ID NO:93, SEQ ID NO:94; their complementary  
30 sequences ; their equivalent sequences, in particular homologous to SEQ ID NO:93, SEQ ID NO:94;

sequences encoding at least part of the peptide sequence defined by SEQ ID NO:95;

sequences encoding at least part of a peptide  
35 sequence equivalent, in particular homologous to SEQ ID NO:95, which is capable of being recognized by sera of



patients infected with the MSRV-1 virus, or in whom the MSRV-1 virus has been reactivated.

10. Nucleic acid probe for the detection of a virus associated with multiple sclerosis or rheumatoid  
5 arthritis, characterized in that it is capable of hybridizing specifically with any fragment according to any one of claim 7 to 9.

11. Probe as claimed in claim 10, consisting of between 10 and 1,000 monomers.

10 12. Primer for the amplification by polymerization of an RNA or a DNA of a viral material associated with multiple sclerosis or rheumatoid arthritis, comprising a nucleotide sequence identical or  
15 equivalent to at least one portion of the nucleotide sequence of a fragment as claimed in any one of claims 7 to 9, in particular a nucleotide sequence displaying, for any succession of at least 10 contiguous monomers, preferably 15 contiguous monomers, more preferably 18  
20 contiguous monomers and even most preferably 20 contiguous monomers, at least 70% homology with at least the said portion of the said fragment.

13. Primer as claimed in Claim 12, comprising a sequence selected from the group consisting of SEQ ID NO: 99 to SEQ ID NO: 111.

25 14. Polypeptide encoded by any open reading frame belonging to a nucleotide fragment as claimed in any one of claims 7 to 9.

15. Polypeptide of claim 14, characterized in that the open reading frame encoding it, is comprised, in  
30 the 5'-3' direction, between nucleotide 18 and nucleotide 2304 of SEQ ID NO:93.

16. Polypeptide according to claim 15, comprising a peptide sequence at least partially identical to SEQ ID NO: 95.

35 17. Polypeptide, comprising a peptide sequence at least partially identical to SEQ ID NO: 96.

18. Polypeptide of claim 17 exhibiting an enzymatic activity consisting of proteolytic activity.

19. Polypeptide, characterized in that the open reading frame encoding it begins, in the 5'-3' direction,  
5 at nucleotide 18 and ends at nucleotide 340 of SEQ ID NO:93.

20. Polypeptide exhibiting an inhibitory activity on the proteolytic activity of polypeptide of claim 18.

10 21. Polypeptide, comprising a peptide sequence identical or equivalent to SEQ ID NO: 97.

22. Polypeptide of claim 21, comprising a peptide sequence identical or equivalent to SEQ ID NO: 98.

15 23. Polypeptide, characterized in that the open reading frame encoding it begins, in the 5'-3' direction, at nucleotide 341 and ends at nucleotide 2304 of SEQ ID NO:93.

24. Polypeptide, characterized in that the open reading frame encoding it begins, in the 5'-3' direction,  
20 at nucleotide 1858 and ends at nucleotide 2304 of SEQ ID NO:93.

25. Polypeptide of claim 21 or 23, exhibiting a reverse transcriptase activity.

26. Polypeptide of claim 22 or 24, exhibiting a  
25 ribonuclease H activity.

27. Polypeptide exhibiting an inhibitory activity on the reverse transcriptase activity of polypeptide of claim 25.

28. Polypeptide having an inhibitory activity  
30 on the ribonuclease H activity of polypeptide of claim 26.

29. Antigenic polypeptide recognized from the sera of patients infected with the MSRV-1 virus, and/or in whom the MSRV-1 virus has been reactivated, characterized in that its peptide sequence is at least partially  
35 identical or is equivalent to a sequence selected from the group consisting of SEQ ID NO:95, and fragments thereof,

in particular SEQ ID NO:96, SEQ ID NO:97 and SEQ ID NO: 98.

30. Mono- or polyclonal antibody directed against the MSRV-1 virus, characterized in that it is  
5 obtained by the immunological reaction of a human or animal body or cells to an immunogenic agent consisting of an antigenic polypeptide of claim 29.

31. Reagent for detection of the MSRV-1 virus, or of an exposure to the said virus, characterized  
10 in that it comprises at least one reactive substance selected from the group consisting of a probe as claimed in claim 10 or 11 ; a polypeptide as claimed in any one of claims 14 to 29 ; or an antibody as claimed in claim 30.

32. Diagnostic, prophylactic or therapeutic  
15 composition, in particular for inhibiting the expression of a virus associated with multiple sclerosis or rheumatoid arthritis, and/or the enzymatic activity of the proteins of said virus, said composition comprising a nucleotide fragment of any one of claims 7 to 9.

20 33. Diagnostic, prophylactic or therapeutic composition comprising a polypeptide of any one of claims 14 to 29, or an antibody of claim 30.

34. Process for detecting a virus associated with multiple sclerosis or rheumatoid arthritis, in a  
25 biological sample, characterized in that an RNA and/or a DNA presumed to belong or originating from said virus, or their complementary RNA and/or DNA, is/are brought into contact with a nucleotide fragment according to any one of claim 7 to 9.

30 35. Process for detecting the presence or exposure to a virus associated with multiple sclerosis or rheumatoid arthritis, in a biological sample, wherein said sample is brought into contact with a polypeptide, according to any one of claim 14 to 29, or an antibody of  
35 claim 30.

1/69

## FIG. 1

Consensus GTTACGGGAT ANCCCTCATC TCTTGGTCA GGTACTGGCC CAAGATCTAG 50  
 Consensus GGCATCTTC AGGTCCAGSN ACTCTGTGCC TTCCAG 85

SEQ ID NO3 (POL MSRV-1B)

Consensus GTTACGGGAT AGCCCCCATC TATTGGGCA GGCATAGCT CAATCTTGA 50  
 Consensus GGCATCTTC ATACTGGAC ATCTGTGCC TTGGT 85

SEQ ID NO4 (POL MSRV-1B)

Consensus GTTACGGGAT AGCCCCCATC TATTGGGCA RGYATTAGC CAAGACTGA 50  
 Consensus GYCACTCTC ATACTGGAC ACTCTGTGCC TTYRG 85

SEQ ID NO5 (POL MSRV-1B)

Consensus GTTACGGGAT AGCTCCATC TATTGGGCT GGCATTAAC CGAGACTTA 50  
 Consensus GGCATCTCT ATACTGGAC ACTCTGTGCC TTGG 85

SEQ ID NO6 (POL MSRV-1B)

Consensus GTGTGGCAC AGGGGTTAR RGATANCYCY CATCTMTTG GYARGYAT  
 Consensus RRCYCRAKAY YTRGYCAVT TCTYAKRYSY RGSNAYTCTB KYCCTTYRGT  
 Consensus ACATGGATGA C

SEQ ID NO7 (POL MSRV-1B)

2/69

FIG. 2

## CONSENSUS A

SEQ ID NO 3

GTTAGGGATAGCCC TCATCTCTTGGTCA GGTACTGGCCCAAGA TCTAGGCCACTTCTC 60  
 V . G . P S S L W S G T G P R S R P L L  
 F R D S P H L F G Q V L A Q D L G H F S  
 L G I A L I S L V R Y W P K I . A T S Q

AGGTCCAGGCACTCT GTTCCTTCAG 85  
 R S R H S V P S  
 G P G T L F L Q  
 V Q A L C S F

## CONSENSUS B

SEQ ID NO 4

GTTCAGGGATAGCCC CCATCTATTTGGCCA GGCCTAGCTCAATA CTTGAGCCAGTTCTC 60  
 V Q G . P P S I W P G T S S I L E P V L  
 F R D S P H L F G Q A L A Q Y L S Q F S  
 S G I A P I Y L A R H . L N T . A S S H

ATACCTGGCACTCT TGTCCTTCGGT 86  
 I P G H S C P S  
 Y L D T L V L R  
 T W T L L S F G

## CONSENSUS C

SEQ ID NO 5

GTTCAGGGATAGCCC CCATCTATTTGGCCA GGCATTAGCCCAAGA CTTGAGTCAATTCTC 60  
 V Q G . P P S I W P G I S P R L E S I L  
 F R D S P H L F G Q A L A Q D L S Q F S  
 S G I A P I Y L A R H . P K T . V N S H

ATACCTGGCACTCT TGTCCTTCAG 85  
 I P G H S C P S  
 Y L D T L V L Q  
 T W T L L S F

## CONSENSUS D

SEQ ID NO 6

GTTCAGGGATAGCTC CCATCTATTTGGCCT GGCATTAACCCGAGA CTTAAGCCAGTTCTC 60  
 V Q G . L P S I W P G I N P R L K P V L  
 F R D S S H L F G L A L T R D L S Q F S  
 S G I A P I Y L A W H . P E T . A S S H

ATACGTGGCACTCT TGTCCTTTGG 85  
 I R G H S C P L  
 Y V D T L V L W  
 T W T L L S F

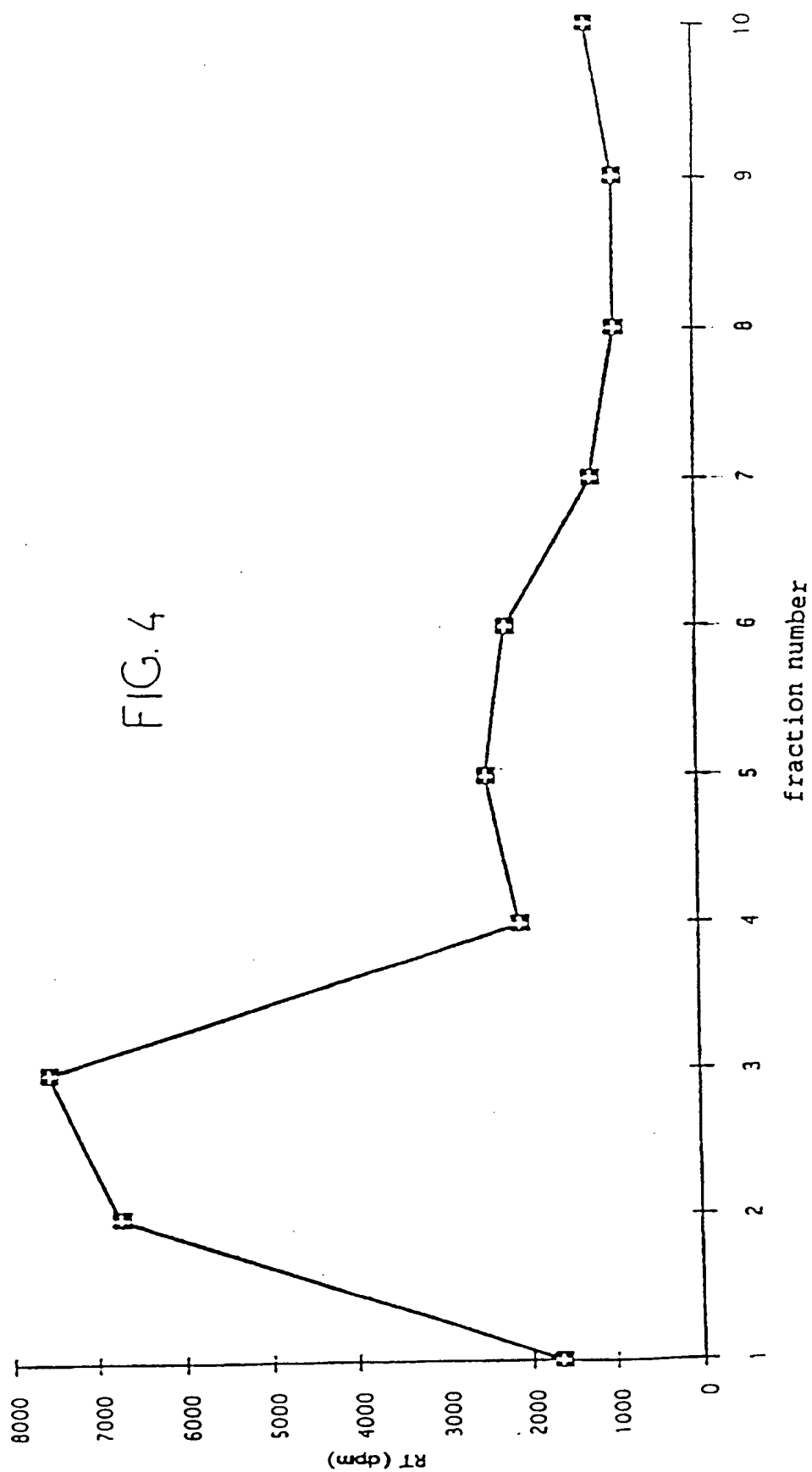
3/69

FIG. 3

Consensus	TTGGATCCAG TGYTGGCCACA GGGCGCTGAA GGCATATGGG TGCAGTTGGC	50
Consensus	GGATGCGGCG TATAGCCCTCT ACGTGGATGA CCTSCGGAAG CTTGAG	96

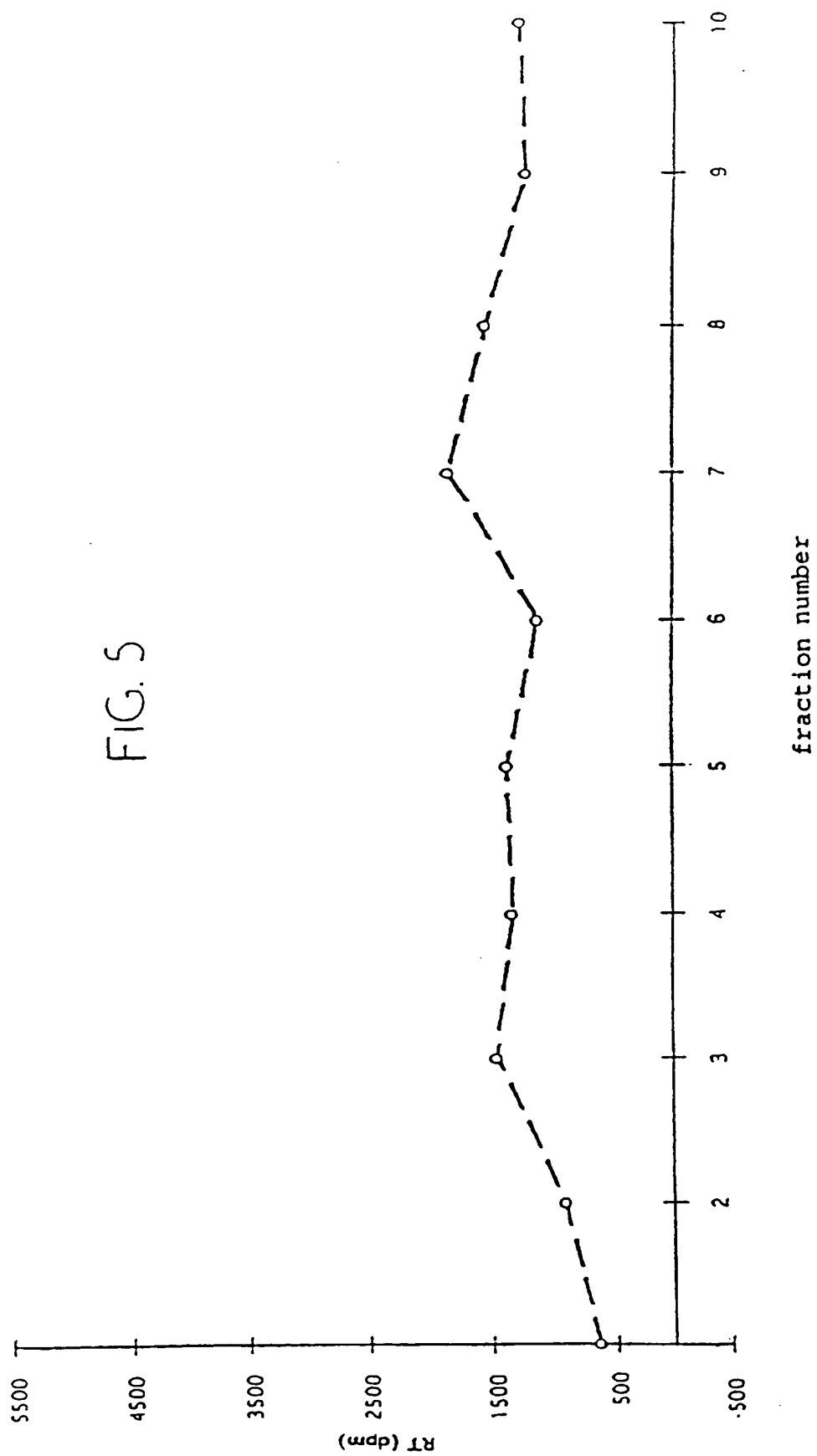
SEQ ID NO 11

4/69



5/69

FIG. 5





6/69

## FIG. 6

CAAGCCACCC AAGAACTCTT AAATTTCTCT ACTACCTGTG GCTACAAGGT	50
TTCCAAACCA AAGGCTCAGC TCTGCTCACA GGAGATTAGA TACTTAGGGT	100
TAAAATTATC CAAAGGCACC AGGGGCTCA GTGAGGAACG TATCCAGCCT	150
ATACTGGGTT ATCCTCATCC CAAAACCTA AAGCAACTAA GAGGGTTCCT	200
TAGCATGATC AGGTTTCTGC CGAAAACAAG ATTCCCAGGT ACAACCAAAA	250
TAGCCAGACC ATTATATACA CTAATTAAGG AAACTCAGAA AGCCAATACC	300
TATTTAGTAA GATGGACACC TAAACAGAAG GCTTTCCAGG CCTTAAAGAA	350
GGCCCTAACC CAAGCCCCAG TGTTCAGCTT GCCAACAGGG CAAGATTTTT	400
CTTTATATGG CACAGAAAAA ACAGGAATCG CTCIAGGAGT CCTTACACAG	450
GTCCGAGGGA TGAGCTTGCA ACCCGTGGCA TACCTGAATA AGGAAATIGA	500
TGTAGTGGCA AAGGTTGGC CTCATNGTTT ATGGGTAATG GNGCAGTAG	550
CAGTCTNAGT ATCTGAAGCA GTTAAAATAA TACAGGGAAG AGATCTTNC	600
GIGIGGACAT CTCATGATGT GAACGGCATA CTCACTGCTA AAGGAGACTT	650
GTGGTGTCA GACAACCAT TACTTAANTA TCAGGCTCTA TTA CTIGAAG	700
AGCCAGTGCT GNGACTGGC ACTTGTGCAA CTCTTAAACC C	741

SEQ ID NO9 (PSJ 17)

7/69

TCAGGGATAGCCCCCATCTATTTGGCCAGGCATTAGCCCAAGACTTGAGTC  
AATTCTCATACCTGGACACTCTTGTCTTCAGTACATGGATGATTTACTTT  
TAGTCGCCCCGTTTCAGAAACCTTGTGCCATCAAGCCACCCAAGAACTCTTAA  
CTTTCCTCACTACCTGTGGCTACAAGGTTTCCAAACCAAAGGCTCGGCTCT  
GCTCACAGGAGATTAGATACTNAGGGCTAAAATTATCCAAAGGCACCAGG  
GCCCTCAGTGAGGAACGTATCCAGCCTATACTGGCTTATCCTCATCCCCAAA  
ACCCTAAAGCAACTAAGAGGGTTCCTTGGCATAACAGGTTTCTGCCGAAA  
ACAGATTCCCAGGTACASCCCAATAGCCAGACCATTATATACACTAATTA  
NGGAAACTCAGAAAGCCAATACCTATTTAGTAAGATGGACACCTACAGAA  
GTGGCTTTCCAGGCCCTAAAGAAGGCCCTAACCCAAGCCCCAGTGTTTCAGC  
TTGCCAACAGGGCAAGATTTTTCTTTATATGCCACAGAAAAAACAGGAAT  
AGCTCTAGGAGTCCTTACGCAGGTCTCAGGGATGAGCTTGCAACCCGTGGT  
ATACCTGAGTAAGGAAATTGATGTAGTGGCAAAGGGTT

SEQ ID NO 8 (MOO3-POO4)

FIG. 7



9/69

10 20 30 40 50 60 70  
CCC TTT GCC ACT ACA TCA ATT TTA GGA GTA AGG AA . CCC AAC GGA CAG TGG AGG TTA GTG CAA GAA CTC AGG  
P F A T T S I L G V R K P N G Q W R L V O E L R>  
TRANSLATION OF MSRV-1 POL \* [A]

80 90 100 110 120 130 140  
ATT ATC AAT CAG GCT GTT GTT OCT CTA TAC CCA OCT GTA OCT AAC OCT TAT ACA GTG CTT TCC CAA ATA CCA  
I I N E A V V P L Y P A V P N P Y T V L S Q I P>  
TRANSLATION OF MSRV-1 POL \* [A]

150 160 170 180 190 200 210  
GAG GAA GCA GAG TGG TTT ACA GTC CTG GAC CTT AAG GAT GGC TTT TTC TGC ATC OCT GTA OCT OCT GAC TCT  
E E A E W F T V L D L K D A F F C I P V R P D S>  
TRANSLATION OF MSRV-1 POL \* [A]

220 230 240 250 260 270 280  
CAA TTC TTG TTT GGC TTT GAA GAT OCT TTG AAC CCA AGG TCT CAA CTC ACC TGG ACT GTT TTA CCC CAA GGG  
Q F L F A F E D P L N P T S Q L T W T V L P Q G>  
TRANSLATION OF MSRV-1 POL \* [A]

290 300 310 320 330 340 350 360  
TTC AGG GAT AGC CCC CAT CTA TTT GGC CAG GCA TTA GGC CAA GAC TTG AGT CAA TTC TCA TAC CTG GAC ACT  
F R D S P H L F G Q A L A Q D L S Q F S Y L D T>  
TRANSLATION OF MSRV-1 POL \* [A]

370 380 390 400 410 420 430  
CTT GTC CTT CAG TAC ATG GAT GAT TTA CTT TTA GTC GGC GGT TCA GAA ACC TTG TGC CAT CAA GGC ACC CAA  
L V L Q Y H D D L L L V A R S E T L C H Q A T Q>  
TRANSLATION OF MSRV-1 POL \* [A]

440 450 460 470 480 490 500  
GAA CTC TTA ACT TTC CTC ACT ACC TGT GGC TAC AAG GTT TCC AAA CCA AAG GCT GGC CTC TCC TCA CAG GAG  
E L L T F L T T C G Y K V S K P K A R L C S Q E>  
TRANSLATION OF MSRV-1 POL \* [A]

510 520 530 540 550 560 570  
ATT AGA TAC TNA GGG CTA AAA TTA TCC AAA GGC ACC AGG GGC CTC AGT GAG GAA GGT ATC CAG OCT ATA CTG  
I R Y X G L K L S K G T R A L S E E R I Q P I L>  
TRANSLATION OF MSRV-1 POL \* [A]

580 590 600 610 620 630 640  
OCT TAT OCT CAT CCC AAA ACC CTA AAG CAA CTA AGA GGC TTC CTT GGC ATA ACA GGT TTC TCC GGA AAA CAG  
A Y P H P K T L K Q L R G F L G I T G F C R K Q>  
TRANSLATION OF MSRV-1 POL \* [A]

650 660 670 680 690 700 710 720  
ATT CCC AGG TAC ASC CCA ATA GGC AGA CCA TTA TAT ACA CTA ATT AAG GAA ACT CAG AAA GGC AAT ACC TAT  
I P R Y X P I A R P L Y T L I X E T Q K A N T Y>  
TRANSLATION OF MSRV-1 POL \* [A]

730 740 750 760 770 780 790  
TTA GTA AGA TGG ACA OCT ACA GAA GTG GCT TTC CAG GGC CTA AAG AAG GGC CTA ACC CAA GGC CCA GTG TTC  
L V R W T P T E V A F Q A L K K A L T Q A P V F>  
TRANSLATION OF MSRV-1 POL \* [A]

800 810 820 830 840 850 860  
AGC TTG CCA ACA GGC CAA GAT TTT TCT TTA TAT GGC ACA GAA AAA ACA GGA ATA OCT CTA GGA GTC CTT AGG  
S L P T G Q D F S L Y A T E K T G I A L G V L T>  
TRANSLATION OF MSRV-1 POL \* [A]

870 880 890 900 910 920 930  
CAG GTC TCA GGC ATG AGC TTG CAA CCC GTG GTA TAC CTG AGT AAG GAA ATT GAT GTA GTG GCA AAG GGT TGG  
Q V S G H S L Q P V V Y L S K E I D V V A K G W>  
TRANSLATION OF MSRV-1 POL \* [A]

940 950 960 970 980 990 1000  
OCT CAT NGT TTA TGG GTA ATG GAG GCA GTA GCA GTC TNA GTA TCT GAA GCA GTT AAA ATA ATA CAG GGA AGA  
P H X L W V H X A V A V X V S E A V K I I Q G R>  
TRANSLATION OF MSRV-1 POL \* [A]

1010 1020 1030 1040 1050 1060 1070 1080  
GAT CTT NCT GTG TGG ACA TCT CAT GAT GTG AAC GGC ATA CTC ACT GCT AAA GGA GAC TTG TGG TTG TCA GAC  
D L X V W T S H D V N G I L T A K G D L S D>  
TRANSLATION OF MSRV-1 POL \* [A]

1090 1100 1110 1120 1130 1140 1150  
AAC CAT TTA CTT AAN TAT CAG GCT CTA TTA CTT GAA GAG CCA GTG CTG NGA CTG GGC ACT TGT GCA ACT CTT  
N H L L X Y O A L L L E E P V L X L R T C A T L>  
TRANSLATION OF MSRV-1 POL \* [A]

AAA CCC  
K P>

FIG. 9

SEQ ID NO 1 (MSRV-1 pol\*)

10/69

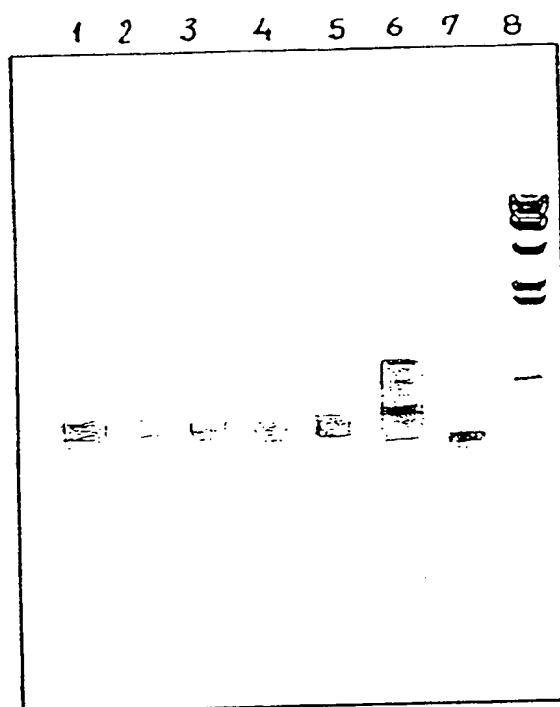
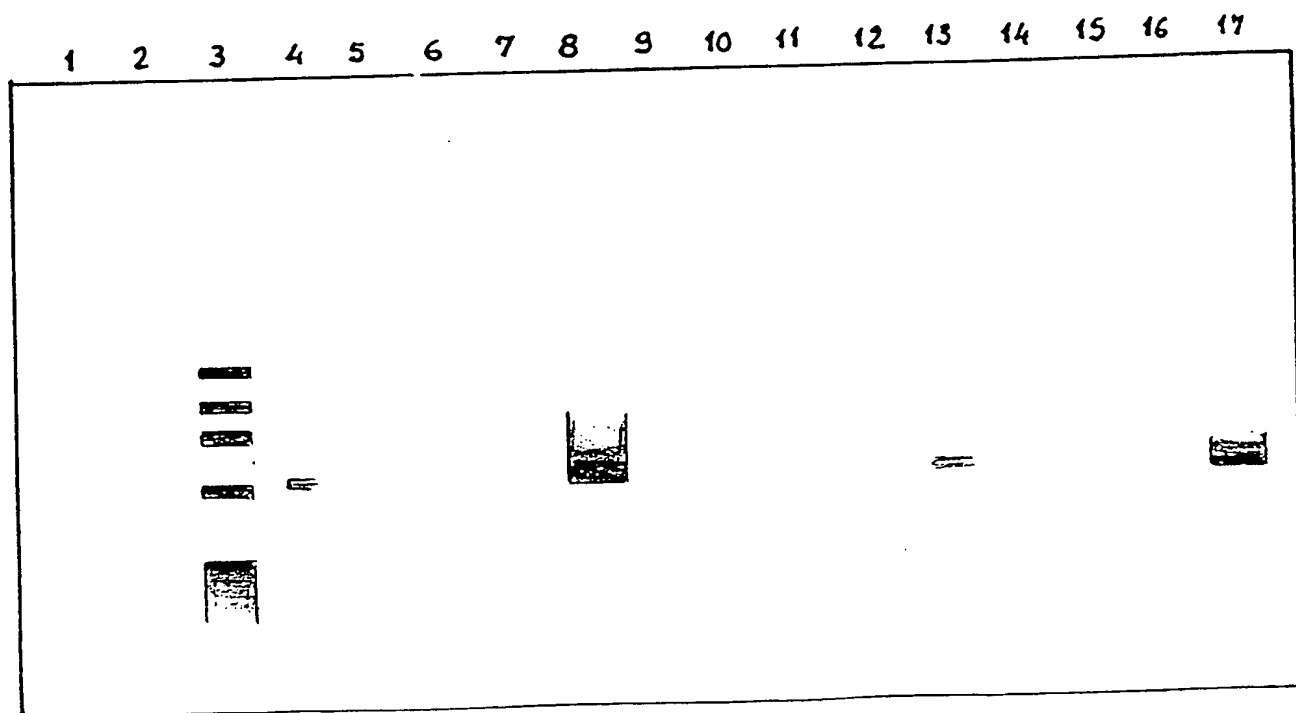


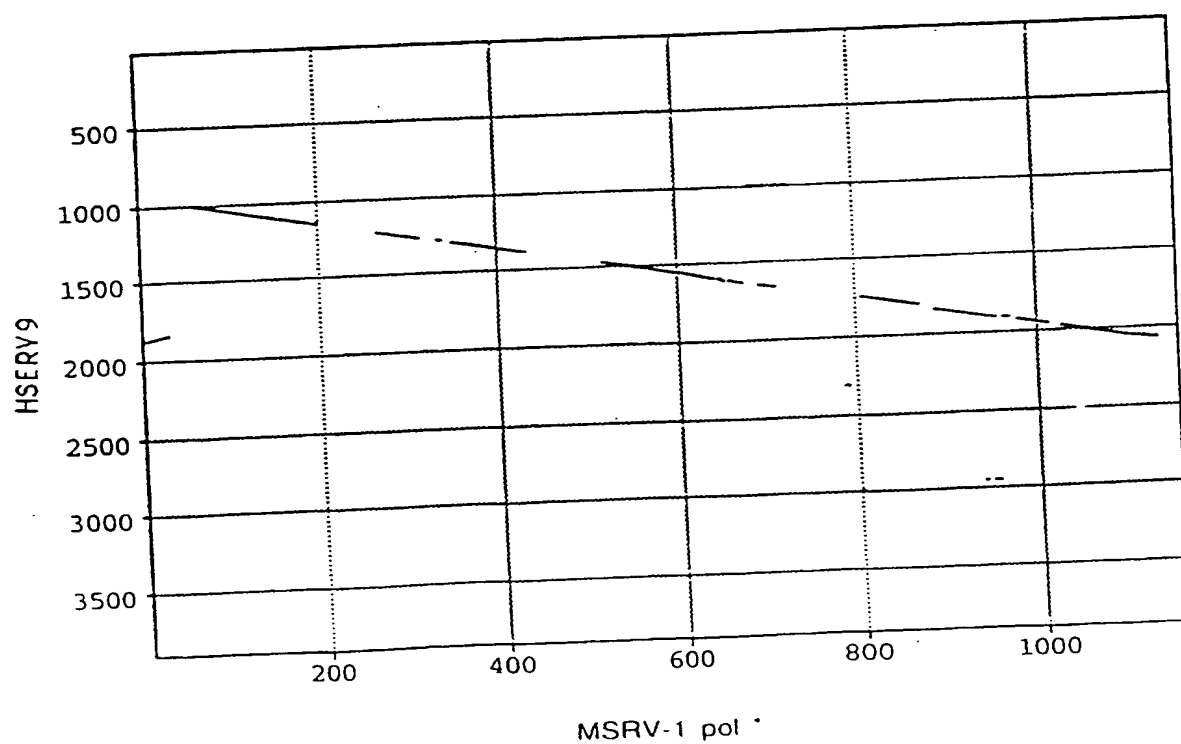
FIG. 10

FIG. 11



11/69

FIG. 12



12/69

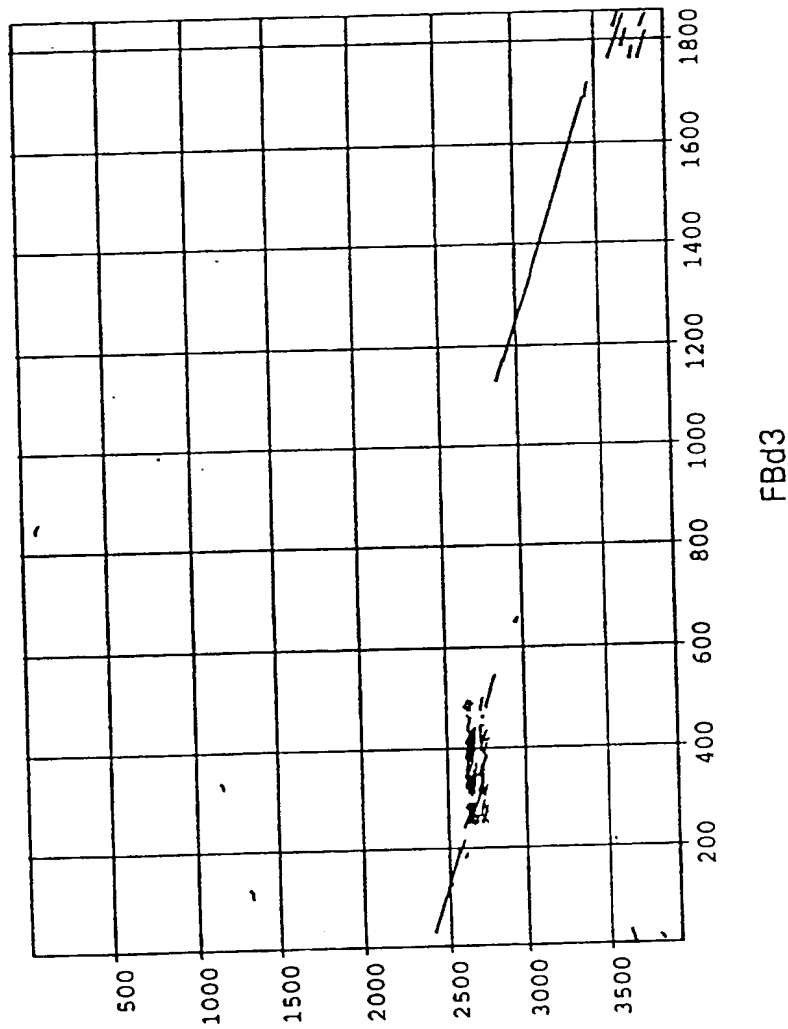
FIG. 13

SEQ ID NO 46 (FBd3)

GTGCTGATTGGTGTATTTACAATCCTTTATCTAATCCGAAATGCCCATGTTG  
CAATATGGAAAGAAAGGGAGTTCCTAACCTCTGGGGGAACCCCCATTAAA  
TACCACAAGTAAATCATGGAGTTATTGCACACAGTGCAAAAACTCAAGGA  
GGTGGAAAGTCTTACACTGCCAAAGCCATCAGAAAAGGGAAGAGGGGAGAA  
GAGCAGCATAAGTGGCTACAGAGGCAAGGAAAGACTAGCAGAAAGGAAA  
GAGAGAAAGAGACAGAAAGTCAGAGAGAGAGAGAGAGGAAGAGACAGAGCA  
CAAAGAGGGAGTCAGAGAGAGAGAGAGAGACAGAGAGTCAGAGAGAAGGAA  
AGAGAGAGAGGAAGAGACAAAGAATGAATCAAACAGAGAGACAGAAAGT  
CAGAGAGAGAGAGAGAGAGAGGAAGAGACAGAGAAAAAGAGGGAGTCAGAA  
AAAGAGAGACCAAAGAAGAAGTCCAAAGAGAAAGAAAGAGAGATGGAAG  
TAGTAAAGGAAAAACAGTGTACCCTATTCTTTTAAAAGCCGGGGTAAATTT  
AAAACCTATAATTGATAACTGAAGGTCTTCTCTGTAACCCTGTAACACTCC  
AATACCACCTTGTGTGCAAGTGTAACAAGGGCGTAGCCCCAAAAGCACTG  
AGGCCACTAACAAACCCATAGCCTTCCTATCAAAATTCCTTAACCCAGCAGG  
TTTCCTAACAGGGGATCTAAATCTTAATTAATTACCATACAATGGTCCAAC  
CAGACTTAGGAGGAATTCCCTTCAGGACGGGAAGATAGATGCTTCCTCCCA  
GGCGATTAAGGGAGAAAGACACAATGGGTATTCAGTAAGTGCCAAGGGGA  
ACACTTGTAGAAGCAAAGTTAGGAAAATTGCCAAATAATTGGTTTGCTCAA  
GAGTTGTTTGCACTCAGCCAAACCTTGAAGTACTTGCAGAATCAGAAAGGA  
GCCATCTATACCAATTCTAAGTTAATATGGACTGAAGGAGGTTTTATTAAT  
ACCAAAGAGAAATTAATAATCCCAAACCTTATAAGGTTTTCAACCAAAGTAA  
AGTTTGCTAAAAGTTAACAGCGTAACATGTATTATCCTACTACCACACACT  
CTCAAAGGATTTCTCAGACAGTTTGCAAGAAATAATGATATCTATCCTTAC  
TCTACAATCCCAAATAGACTCTTTGGCAGCAGTGACTCTCCAAAACCGTCA  
AGGCCTAGACCTCCTCACTGCTGAGAAAGGAGGACTCTGCACCTTCTTAAG  
GGAAGAGTGTTGTCTTTACACTAACCAGTCAGGGATAGTATGAGATGCTGC  
CCGGCATTTACAGAAAAAGGCTTCTGAAATCAGACAACGCCTTTCAAATTC  
CTATACCAACCTCTGGAGTTGGGCAACATGGTTTCTTCCCTTTCTATGTCCC  
ATGGCTGCCATCTTGCTATTACTCGCCTTTGGGCCCTGTATTTTTAACCTCC  
TTGTCAAATTTGTTTCTTCTAGGATCGAGGCCATCAAGCTACAGATGGTCTT  
ACAAATGGAACCCCAAATGAGCTCAACTATCAACTTCTACTGAGGACCCCT  
AGACCAACCCCTGGCCCTTTCACTGGCCTAAAGAGTTCCCGTCTGGAGGA  
CACTACCACTGCAGGGCCCCATCTTTGCCCTATCCAGAAGGAAGTAGCTA  
GAGCAGTCATTGCCCAATTCCCAAGAGCAGCTGGGGTGTCCCGTTTAGAGT  
GGGGATTGAGAGGTGAAGCCAGCTGGACTTCTGGGTGCGGTGGGGACTTG  
GAGAACTTTTGTGTCTAGCTAAAGGATTGTAAATGCAACAATCAGTGCTCT  
GTGTCTAGCTAAAGGATTGTAAATACCAATCAGCAC

13/69

FIG. 14



HSERV9



14/69

## FIG. 15

SEQ ID NO 51 ( t pol)

GGCTGCTAAAGGAGACTTGTGGTTGTCAGACAATCGCCTACTTAGGTACCA  
GGCCTTATTACTTGAGGGACTGGTGCTTCAGATGCGCACTTGTGCAGCTCT  
TAACCCAAACTTATGCTGCCCAGAAGGATCTTTTAGAGGTCCCCTTAGCCA  
ACCCTGACCTCAACCTATATATATACTGATGGAAGTTCGTTTGTAGAAAAG  
GGATTACAAAGGGNAGGATATNCCATAGGTAGTGATAAAGCAGTACTTG  
AAAGTAAGCCTCTTCCCCCAGGGACCAGCGCCCCCGTTAGCAGAACTAGT  
GGCACTGACCCCGAGCCTTAGAACTTGGAAAGGGAGGAGGATAAATGTGT  
ATACAGATAGCAAGTATGCTTATCTAATCCGAAATGCCCATGTTG

15/69

SEQ ID NO 52 (JLBc1)

TCAGGGATAGCCCCCATCTATTTGGTCAGGCACCTGGCCCAAGATCTAGGGA  
CATGCCACTTTTAAAGAGCCATTTCTCAAGTCCAGGTAAGTCTGGTCCCTCGGT  
ATGTGGATGATTTACTTTTGGCTACCAAGTTCAGTAGCCTCATGCCAGCAGG  
CTACTCTAGATCTCTTGAACCTTTCTAGCTAATCAAGGGTACAAGGCATCTA  
GGTTGAAGGCCCAGCTTTGCCTACAGCAGGTCAAATATCTAGGCCTAATCT  
TAGCCAGAGGGGACCAGGGGCACTCAGCAAGGAACAAATACAGCCTATACTG  
GCTTATCCTCACCCCTAAGACATTAACACAGTTGCGGGGGTTCCTTGAATC  
ACTGGCTTTTTTGGTGACTATGGATTCCAGATACAGCAAGATTGGCAGGCC  
CCTCTATACTGTAATCAAGGAGACTCACGAGGGCAAGTACTCATCTAGTAG  
AATGGGAAGTACAGGACAGAAACAGCCTTCAAAACCTTAAAGCAGGCCCTA  
GTACAATCTCCAGCTTTAAGCCTTCCCACAGGACAAAACCTTCTCTTTATAC  
ATCACAGAGAGGGCAGAGATAGCTCTTGGTGTCCTTATTCAGACTCATGGG  
ACTACCCCAACAACCAAGTGGCACACCTAAGTAAGGAAATTGATGTAGTAGC  
AAAAGGCTGGCCTCACTGTTTATGGGTAGCTGTGGTGGTGTCTTAGT  
GTCAGAAGCTATCAAAATAATACAAGGAAAGGATCTCACTGTCTGGACTA  
CTCATGATGTAATGGCATACTAGGTGCCAAAAGAAAGTTTATGGGTATCAGA  
CAACCACCTGCTTAGATAACAGGGGACTACTCCTGGAGGATTGGGCTTCAAG  
TGCGTTTTTTGTGGCCTCAACCCTGCCACTTTTCTCCAGAGGATGGAGAG  
CCGCTTGAGCATGCTTGCCAACAGGTTGTAGGCCAGAATTATTCACCCGA  
GATGATCTCTTAGAGTACCCTTAGCTAATCCTGACCTTAACCTATATACCA  
ATGGAAGTTTCAATTTGTGGAAAACGGGATATGAAGGGCAGGTTATGTCATAG  
TTAGTGATGTAATCATACTTGCAAGTAAGCCTCTTACCCCAAGGGGCCAGCA  
CTCAGTTAGCAGAACTAGTCACACTTACCTTAACCTTAGAACTGGGAAAGG  
GAAAAAGAATAAATATGTATACAGATAGTAAGTATGCTTATCTAATCCTAC  
ATGCCCATGCTGCAATATGGAAGGAAAGGGAGTTCCTAACCCTTGGGGGA  
ACCCCATTAATAATACCACAAGGYAAATCATGGAGTTATTGCACGCAGTGC  
AAAAACTCAAGGAGGTGGCAGTCTTACACTGCCGAAGCYATCAAAAAGGG  
GAAGGAGAGGGGAGAACAGCAGCATAAGTGGTTGGCAGAGGCAGTGAAG  
GACCAGCAGAGAGAAGGAGAGAGACAACGTCAACGACAGAAGGAAAGAA  
GAGGAGGAGACAGAGAGGAAGAGACAGAGAGACAGTTAGTCCAAGAGAG  
AGACAGAGAGAGGAAGAGACAGACAGAAAGTCCAAGAGAGAAGGAAAGA  
GAGGAAGAGACCAAGGAGTCCNAGAGAGAGAAAGAGATAGAAGTAGTAA  
AGAAAAAACATTGTACCCTATTCTTTAAAGCCGGGGTATATTTAAACC  
TATAATTGATAATTGAGTTCTTGCACCCTCCTCCAGGGGATYGCTGGGAGG  
AAACCCTCAACCGATATGTGAAAATTGTGGGTCGTCCCTATGTCTCAATTA  
CCAGCCAATACCCCTTGTTTTTAGTGTGAACGAGGGGTGTAGAGCGCAGAC  
AGGGAGACCTCTGACAATCCATACCCTTCTATCCAAAATCCTTAACCCAG  
CAGGTTTTCTAAAAGGGGATCTAAATCTTAATTAATTACCATACAAAGGTC  
AAACCAGATCTAGGAGGAACTTCTTTCAGGACAGGATGATAGATGGTTCCT  
CCCAGGCGATTAAAGAAAATAAAAAGACACATGGGCAGCCAGTAAGTGAT  
AAGGGAACACTAGTAGAAGCAGTTAGGAGAAGTTGCCTAATAATTGGTCT  
ACTCCAAATGTGTGAGTTGTTTCGCACTCAGCCCAAATCTTAAAGTACTTAC  
AGAATTAGGGAGGAGCCATTTACACCAATTCTAAGTTAATATGGACTGGAT  
GAGGTTTTATTAATAGCGAAGGAGAATTAAATCCTAAACTNACAAGGTTTT  
CAACTAAAGTAAATTTTACTAAAAGCTAACAGTGTAACATGCATTATCCTA  
CTACAACACACTCTCANAGGATTCTCAGACAGTTTACAAGAAATAACAA  
AATCTATCTGGTAAGGATAGTAACATAATCCCAAATACATTCTTTGGCAG  
CAGTGACTCTC

FIG. 16

16/69

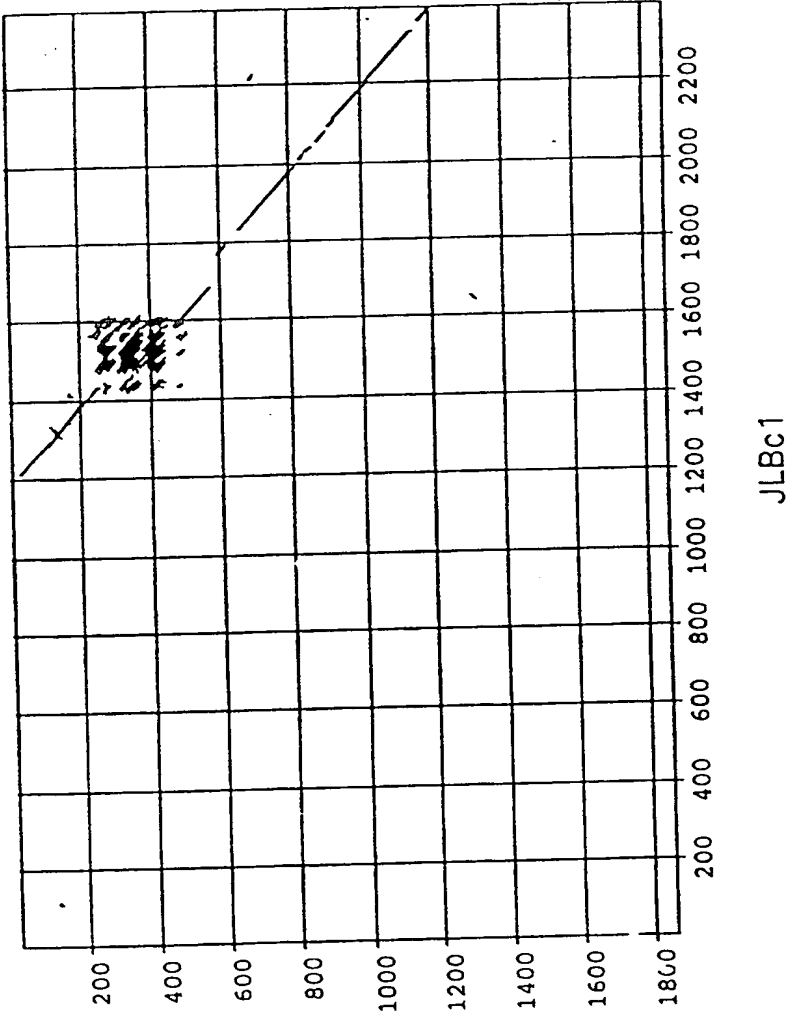
SEQ ID NO 53 (JLBc2)

TCAGGGATAGCCCCC.ATCTATTTGATCAGGCACTAGCCCCAAGATCTAGGCC  
ACTTCTGAAGTCCAGGCATTCTAGTCCTTCAGTATGTGGATGATTTACTTTT  
GGCTACCAGTTTGGGAAGCCTCATGCCAGCAGGCTACTTGAGATCTCTTGAA  
CTTTCTAGCTAATCAAGGGTGTATGGCATCTAAATTGAAAGTCCAGCTCTG  
CCTACAACAAGTCAAATATCTAGGCCTAATCTTAGATAGAAGAACCAGGG  
CCCTCAGCAAGGAATGAATAAAGCCTATGCTGGCTTATCGGCACCCTAAGA  
CATTAACAATTGTGGGGGTTCCTTGGGAATCACTGGCTTTTGCCGACTAT  
GGATCCCTGGATAGAGTGAGATAGCCAGGCCCCCTCTATTACTCTTATCAA  
GGAGACCCAGAGAGGGCAAATACTTATCTAGTATTATGGGNACCAGAGGCAG  
AAAAAGCCTTCCAAACCTTAAAGGAGACCCTAGTACAAGCTCCAGCTTTAA  
GCCTTCCCACAGGACAAANCTTCTCTTTATATGTCACAGAGAGAGCAGGAA  
TAGCTCCTGGAGTCCTTACTCAGACTTTTGGACGACCCACAGGCCAGTGGC  
RTACCTAAGTAAGGAAATTGATGTAGTAGCAAAAGGCTGGCCTCACTGTTT  
ATGGGTAGTTGCGGCTGTGGCAGTCTTACTGTCAAAGGCTATCAAAATAAT  
ACAAGGAAAGGATTTCACTATCTGGACTACTCATGAGGAAAATGGCATATT  
AGGTGCCAAAGGAAGTTTTTGGCTATCAGACAACCACCTGCTCAGATTCCA  
GGCACTACTGATTGAGAGACCAGTGCTTTAAATATGTATGTGTGTGTGG  
CCCTCAACCCTGCCACTGTTCTCCAGAAGATGGAGAACCAATGAAGCATT  
ACTGTCAACAAATTAGAGTCCAGAGTTATGCTGCCTGAGAGGATCTCTTAG  
AAGTCCCCTTAGCTAATCCTGACCTAACCTATATGCTGATGGAAGTTCAC  
TTGTGGAGAATGGGATACGAAAAGCACATTATGCCATAGTTAGTGAGGTA  
ACAGTACTTGAAAGTAAGCCTATTCCCCCATGGACCAGAGCCCAGTTAGCA  
GAACTAGTGGCACTTACCCAAGCCTTAGAACTAGGAAAGGGAAAAATAAT  
AAATGTGTATACAGATAGCAAGTATGCTTATCTAATCCTACATGCCCATGC  
TGCAGTATGGAAAGAAAGGGAGTTCCTAACCTCTGGGGGAACCCCCATTA  
AATACCACAAGGCAAATCATGGAGTTATTGCATGTAGTGCAAAACCTCAA  
GTAGGTGGCAGTTTTACACTGCCTGAAGCTATGGGGAAGGAGAGAGGAGA  
ACAGCAGCATAAGTGGCTAGCAGAGGCAGCGAAAGACTAGCAGAGAGGA  
GAGGTAGGGGAAAGACAGAAAAGTCAAAGAAAAGAGTCAAAGACAGACA  
GAGAAAGAGACAGAGGGAGCCAGAGAGAAAAGAAAAGAGAGAACGAAAGA  
GACAGAATGTCAAAGAACAGAAAGAGAGAGGCAGCGCCAGAAGAGTTAAG  
AAAGTGAGAAAGAGAGATGGAAATAGTAAAGAAAAAACAGTGTACCCTAT  
TCCTTTAAAAGCCAGGGTAAATTTAAAACGTATAATTTTATAATTGGAAGG  
TCTTCTCCATAACCCTATAACATTAAATACCACCTTGTTGTCAAGTGAAAC  
AAGAGCATAGCCCCAAAAGCACTGAGGCCACTGACAACCCATAGCCTTCCT  
ATCAAAAATCCTTAACTCTGCAGGTTTCCTAACAGGGGATCTAAATCTCAA  
CTAATCACCATACAATGGTCCGACCAGACCTAGGAGCGACTCCCCTCAGG  
ACAGAAGGATGGATGGTTCCTCCCAGGCCATTAAGGGAAAGAGACACAAT  
GGGTATTAGTAAGTGATAAGGGAACTCTTGTAGAAGCAGTTAGGAAGATT  
GCCTAATATTTGGTCTGCTCAAATGTGCCAGCTGTTTGCCTCAGCTAAAC  
CTTAAATTACTTACAGAATTAGGAAGGAGCCATCTATAACCAATTCTGAGTT  
AATATGAGCTGAACAAGTTCTTATTAATAGCAAAGAATCATTGAAATCTCA  
AACTTGCAAAGTTTTCAACAAAAGTAAAGTTTGCTGAAAGTTAGCAGTGTA  
ACATGTATTATCCTAACTTCTAATCTTGTGGAAATCAGACCCTATCAGTGC  
CCCTCAAAGCTGAAGTCCATCAGCATATGGCCATACAACTAATACCCCTAT  
TTATAGGGTTAGGAATGGCCACTGCTACAGGAATGGGAGTAACAGGTTTAT  
CTACTTCATTATCCTATTACCACACACTCTTAAAGGATTTCTCAGACAGTTT  
ACAAGAAATAACAAAATCTATCCTTACTCTNTARTCCCAAATAGRTTCTTT  
GGCAGCAGTGACTCTC

FIG. 17

17/69

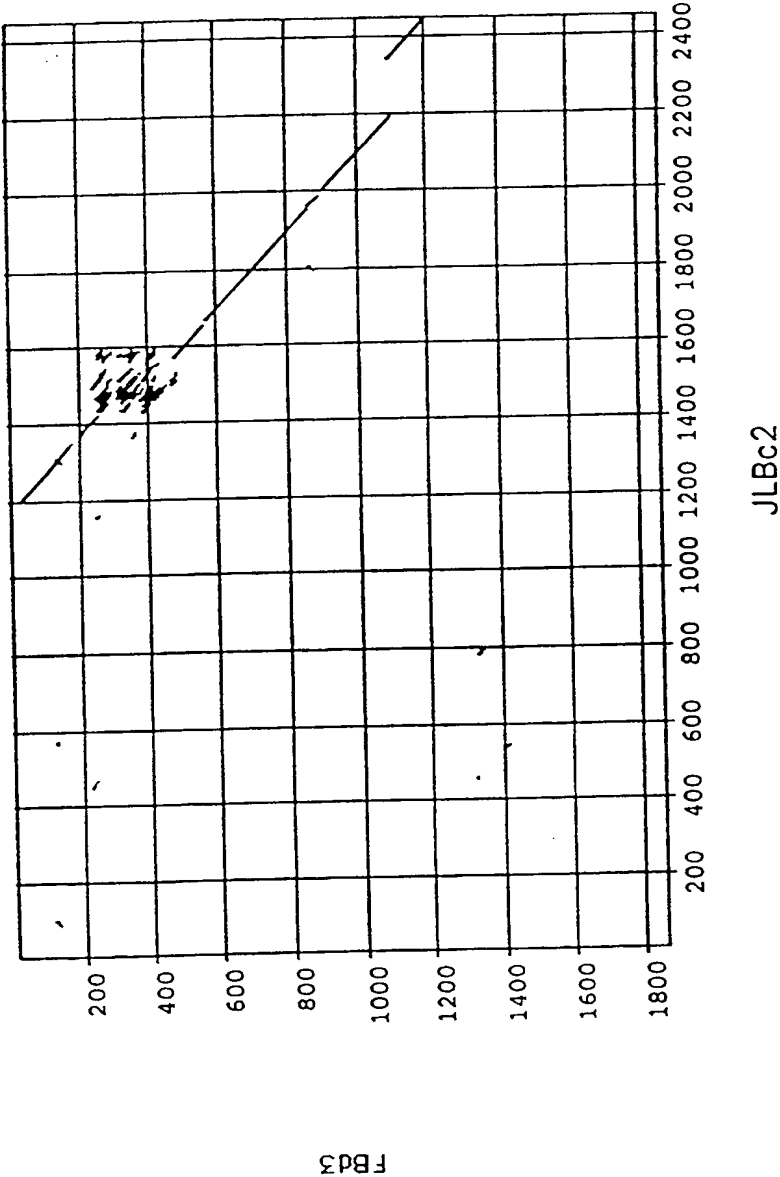
FIG 18



FBd3

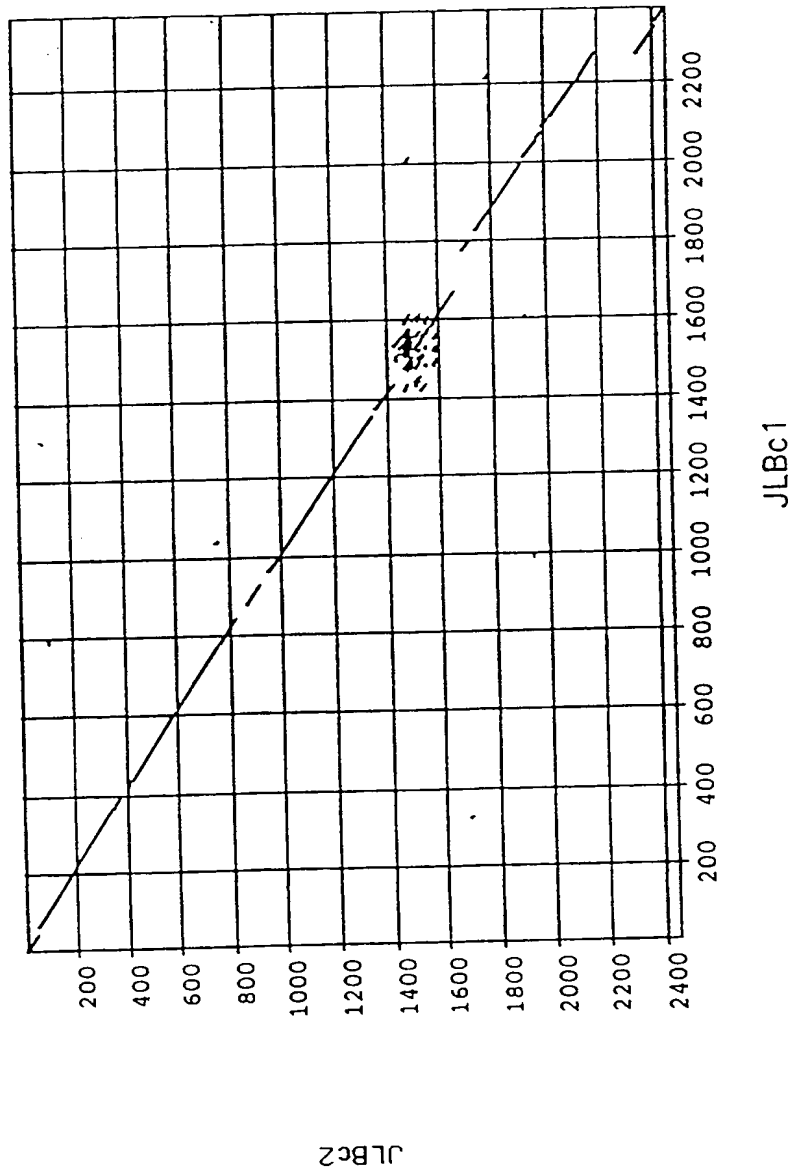
18/69

FIG 19



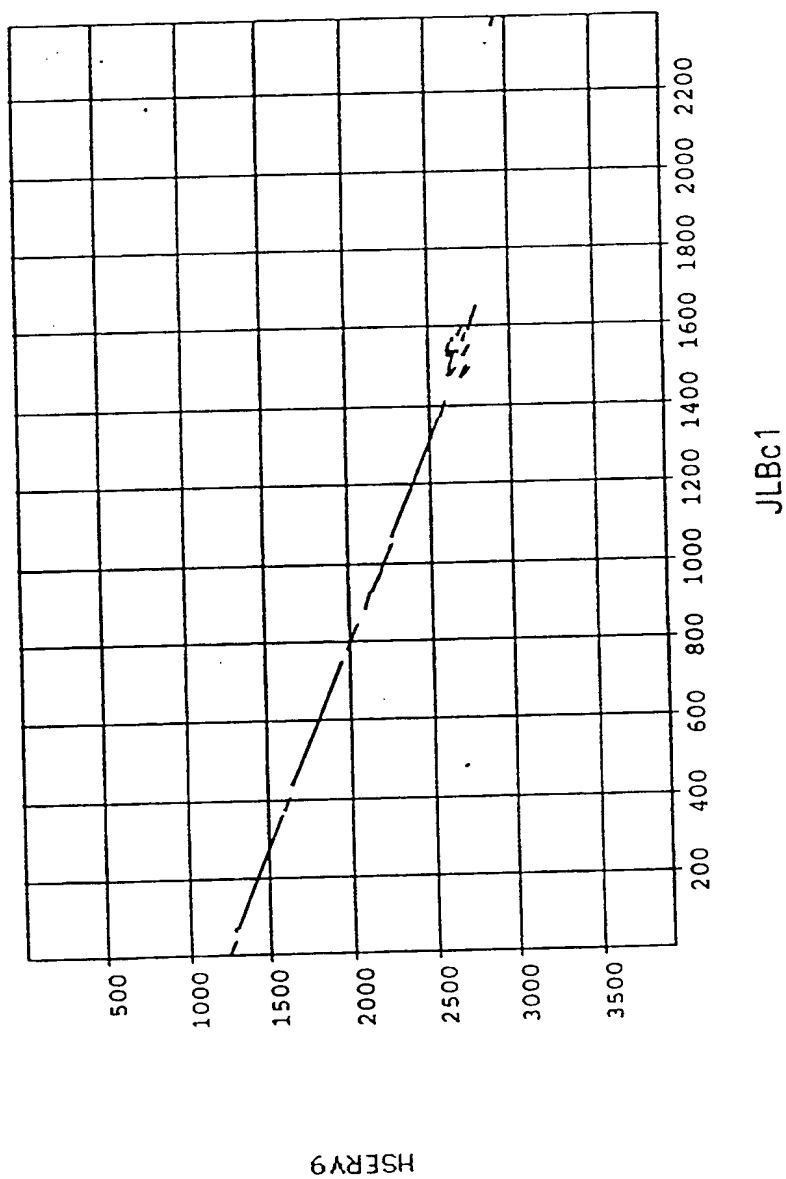
19/69

FIG 20



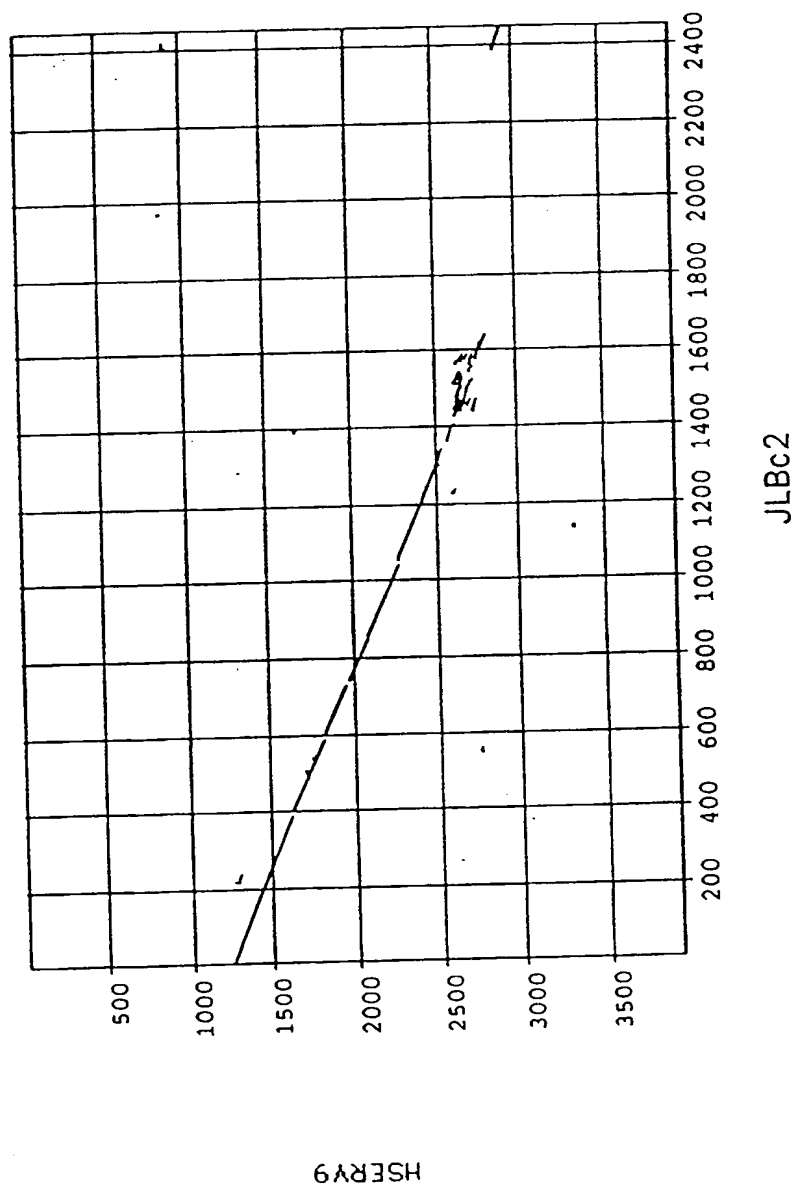
20/69

FIG 21



21/69

FIG 22





22/69

```
1 TTCCTGAGTT CTTGCACTAA CCTCAAATGA GAGAAGTGCC GCCATAACTG CAACCCAAGA
61 GTTTGGCGAT CCCTGGTATC TCAGTCAGGT CAATGACAGG ATGACAACAG AGGAAAGATA
121 ATGATTCCCC ACAGGCCAGC AGGCAGTTCC CAGTGTAGAC CCTCATTAGG ACACAGAATC
181 AGAACATGGA GATTGGTGCC GCAGACATTT GCTAACTTGC GTGCTAGAAG GACTAAGGAA
241 AACTAGGAAG ATATGAATTA TTCAATGATG TCCACTATAA CACAGGGGAA AGGAAGAAAA
301 TCCTACTGCC TTTCTGGAGA GACTAAGGGA GGCATTGAGG AAGCATACCA GGCAAGTGGA
361 CATTTGGAGGC TCTGGAAAAG GGAAAAGTTG GGAAAAGTAT ATGTCTAATA GGGCTTGCTT
421 CCAGTGTGGT CTACAAGGAC ACTTTAAAAA AGATTGTCCA ATAGAAATAA GCCACCACCT
481 CGTCCATGCC CCTTATGTCA AGGGAATCAC TGGAAGGCCC ACTGCCCCAG GGGATGAAGG
541 TCCTCTGAGT CAGAAGCCAC TAACCAGATG ATCCAGCAGC AGGACTGAGG GTGCCCCGGG
601 CAAGCGCCAG CCCATGCCAT CACCCTCACA GAGCCCCAGG TATGCTTGAC CATTGAGGGT
661 CAGAAGGGTA CTGTCTCCTG GACACTGGCG GGCCTTCTCA GTCTTACTTT CCTGTCCTGG
721 ACAACTGTCC TCCAGATCTG TCACTGTCCG AGGGGTCTTA GGACAGCCAG TCACTAGATA
781 CTTCTCCCAG CCACTAAGTT GTGACTGGGG AACTTTACTC TTCCACATGC TTTTCTAATT
841 ATGCCTGAAA GCCCCACTCT CTTGTTAGGG GAGAGACATT CTAGCAAAAG CAGGGGCCAT
901 TATACATGTG AATATAGGAG AAGGAACAAC TGTTTGTGTG CCCCTGCTTG AGGAAGGAAT
961 TAATCCTGAA GTCCGGGCAA CAGAAGGACA ATATGGACAA GCAAAGAATG CCCGTCCTGT
1021 TCAAGTTAAA CTAAAGGATT CCACCTCCTT TCCCTACCAA AGGCAGTACC CCCTCAGACC
1081 CGAGACCCAA CAAGAACTCC AAAAGATTGT AAAGGACCTA AAAGCCCAAG GCCTAGTAAA
1141 ACCAAGCAAT AGCCCTTGCA AGACTCCAAT TTTAGGAGTA AGGAAACCCA ACGGAC
```

SEQ ID NO 56 (GM3)

FIG. 23

23/69

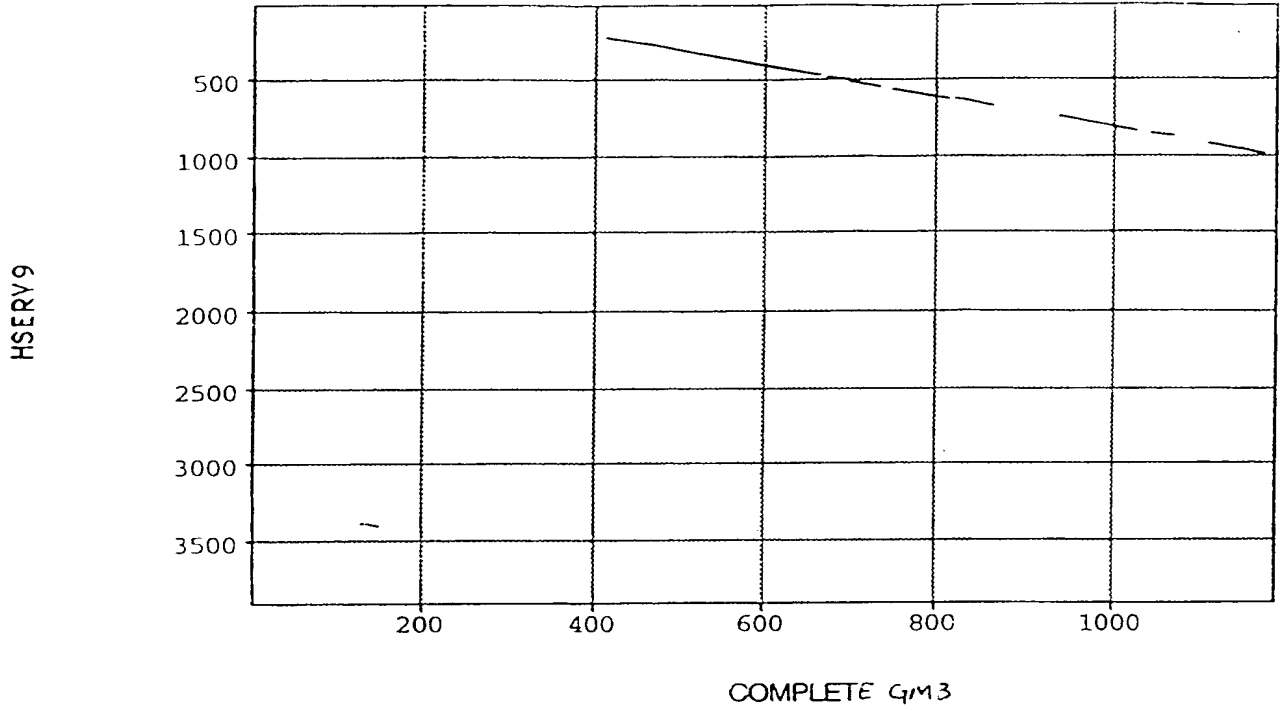
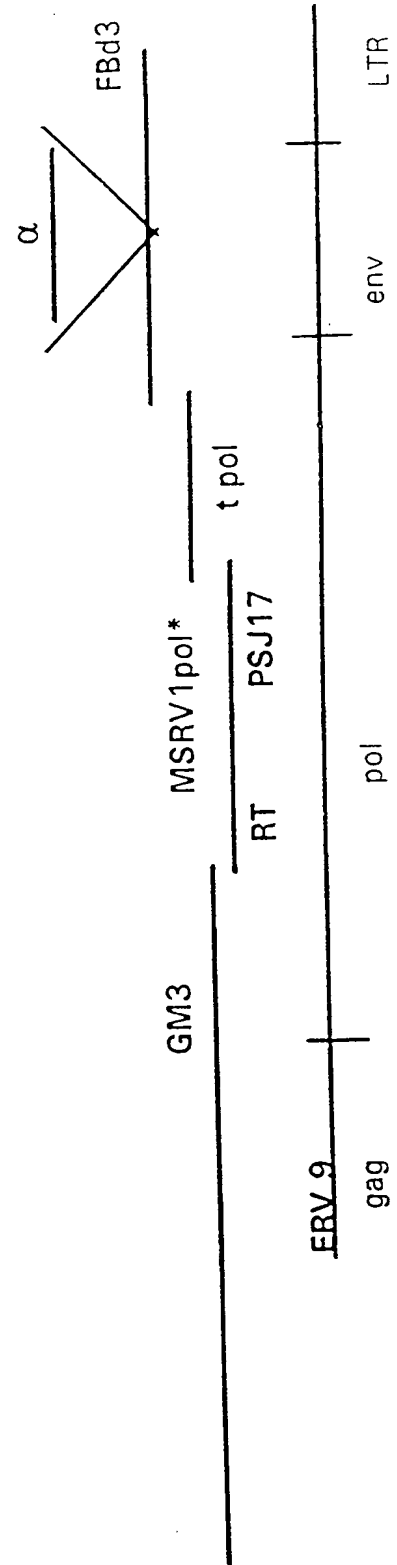
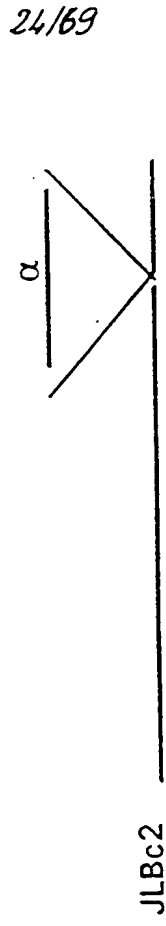
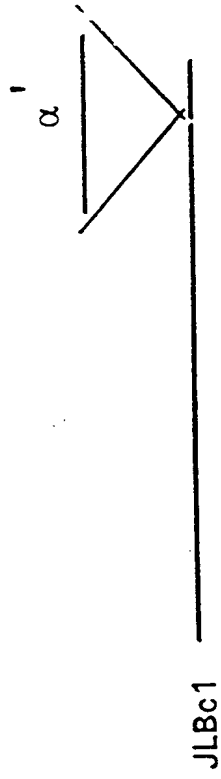


FIG. 24

FIG. 25



25/69

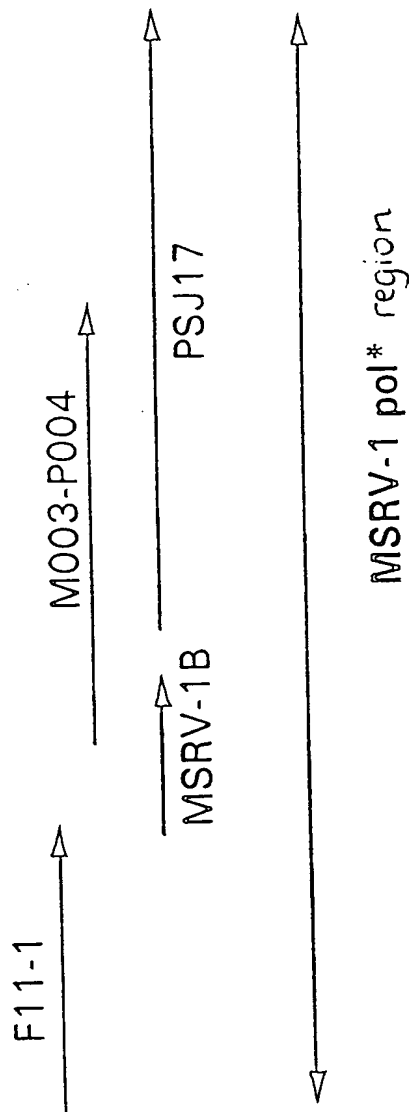


FIG. 26

26/69  
FIG. 27a

[illegible]

BNSDOCID: <WO 9823755A1>

27/69

FIG. 27b

SEQ ID NO 57 (POL)

1170 TCA CAG GAG ATT AGA TAC TNA GCG CTA AAA TTA TCC AAA GCG ACC AGG GCG CTC AGT GAG GAA CGT ATC CAG CCT ATA CTG CCT TAT CCT  
 S Q E I R Y X G L K L S K G T R A L S E E R I Q P I L A Y P  
 1260 CAT CCC AAA ACC CTA AAG CAA CTA AGA GCG TTC CTT GCG ATA ACA GGT TTC TGC CGA AAA CAG ATT CCC AGG TAC ASC CCA ATA GCG AGA  
 H P K T L K Q L R G F L G I T G F C R K Q I P R Y X P I A R  
 1350 CCA TTA TAT ACA ATT ANG GAA ACT CAG AAA GCG ACC AAT ACC TAT TTA GTA AGA TGG ACA CCT ACA GAA GTG GCT TTC CAG GCG CTA AAG  
 P L Y T L I X E T Q K A N T Y L V R W T P T E V A F Q A L K  
 1440 AAG GCG CTA ACC CAA GCG CTA GTC TTC AGC TTC CCA ACA GCG CAA GAT TTT TCT TTA TAT GCG ACA GAA AAA ACA GGA ATA GCT CTA GGA  
 K A L T Q A P V F S L P T G Q D F S L Y A T E K T G I A L G  
 1530 GTC CTT ACG CAG GTC TCA GCG ATG AGC TTC CAA CCC GTG GTA TAC CTG AGT AAG GAA ATT GAT GTA GTG GCA AAG GGT TGG CCT CAT NST  
 V L T Q V S G M S L Q P V Y L S K E I D V V A K G W P H X  
 1620 TTA TGG GTA ATG GAG CCA GTC TNA GTA TCT GAA GCA GGT AAA ATA ATA CAG CGA AGA GAT CTT NCT GTG TGG ACA TCT CAT GAT  
 L W V M X A V A V X V S E A V K I I Q G R D L X V W T S H D  
 1710 GTG AAC GCG ATA CTC ACT CCT AAA GGA GAC TTG TGG TCG AAC CAT TTA CTT AAN TAT CAG GCT CTA TTA CTT GAA GAG CCA GTG  
 V N G I L T A K G D L W L S D N H L L X Y Q A L L L E E P V  
 1800 CTG NGA CTG CCG ACT TGT CCA ACT CTT AAA CCC AAA CTT ATG CTG CCC AGA AGG ATC TTT NTA GAG GTC CCC TTA GCG AAC CCT GAC CTC  
 L X L R T C A T L K P K L M L P R R I F X E V P L A N P D L  
 1890 AAC TAT ATA TAT ACT GAT GGA AGT TCG TTT GTA GAA AAG GGA TTA CAA ACG CNA CGA TAT NCC ATA GGT GGT AGT GAT AAA GGA GTA CTT  
 N Y I Y T D G S S F V E K G L Q R X G Y X I G V S D K A V L  
 1980 GAA ACT AAG CCT CTT CCC CAG GGA CCA GCG CCC TTA GCA CTA GTG CCA ACC CCG CGA GCG TTA GAA CTT TGG AAA GCG  
 E S K P L P P Q G P A P L A E L V A L T P R A L E L W K G  
 2070 ACG ACG ATA AAT GTG TAT ACA GAT ACG AAG TAT GCT TAT CTA ATC CCA AAT GCG CAT GTT TAT CTA ATC CGA AAT GCG CAT GGT CCA  
 R I N V Y T D S K Y A Y L I R N A H V V Y L I R N A H V A  
 2160 ATA TGG AAA GAA ACG GAG TTC CTA ACC TCT GCG GGA ACC CTT AAA TAC CAC AAG TTA ATC ATG GAG TTA TTG CAC ACA GTG CAA AAA  
 I W K E R E F L T S G G T P I K Y H K L I M E L L H T V Q K

SUBSTITUTE SHEET (RULE 26)

28/69

SEQ ID NO 57 (POL)

FIG. 27c

CTC AAG GAG GTG GAA GTC TTA CAC TCC CAA ACC CAT CAG AAA ACG GAA ACG CGA GAG CAG CAT AAG TGG CTA CAG AGC CAA GGA AAG 2250  
L K E V E V L H C Q S H Q K R E R G E E Q H K W L Q R Q G K  
ACT AGC AGA AAG GAA AGA GAG AAA CAG ACA GAA AGT CAG AGA GAG GAA GAG ACA GAG CAC AAA GAG CGA GTC AGA GAG AGA GAG 2340  
T S R K E R E K E T E S Q R E R E E T E H K E G V R E R E  
AGA CAG AGA GTC AGA CAG AAG GAA AGA GAG AGA CGA GGA CAC AAA GAA TGA 2391  
R Q R V R E K E R E R G R D K E .

*29/69*

FIG. 28

GATGCCTTTTTCTGCATCCCTGTACGTCCTGACTCTCAATTCTTGTTTGCCTTTGAAG  
ATCCTTTGAACCCAACGTCTCAACTCACCTGGACTGTTTTACCCCAAGGGTTCAGGGA  
TAGCCCCATCTATTTGGCCAGGCATTAGCCCAAGATGCCTTTTGCATCCCTGTACGTG  
ACTCTCAATTCTTGTTTGCCTTTGCCTTTGAAGATGCTTTGAACCCAACGTCTCAACT  
CACCTGGACTGTTTTACGCCAAGGGTTCAGGGATAGCCCCCATCTATTTGGC  
CAGGCATTAGCCCAA

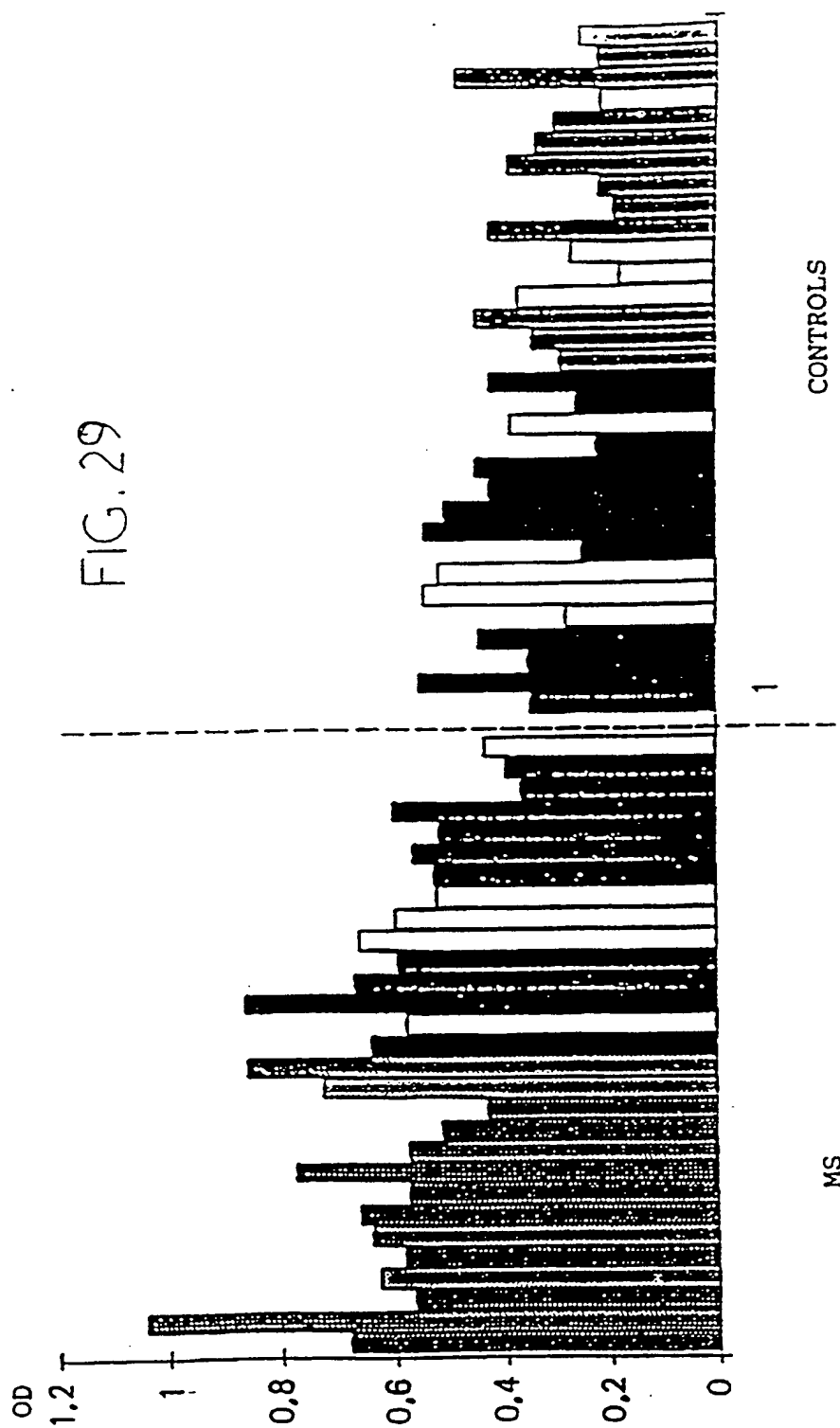
SEQ ID NO 40

Asp-Ala-Phe-Phe-Cys-Ile-Pro-Val-Arg-Pro-Asp-Ser-Gln-Phe-  
Leu-Phe-Ala-Phe-Glu-Asp-Pro-Leu-Asn-Pro-Thr-Ser-Gln-Leu-  
Thr-Trp-Thr-Val-Leu-Pro-Gln-Gly-Phe-Arg-Asp-Ser-Pro-His-  
Leu-Phe-Gly-Gln-Ala-Leu-Ala-Gln

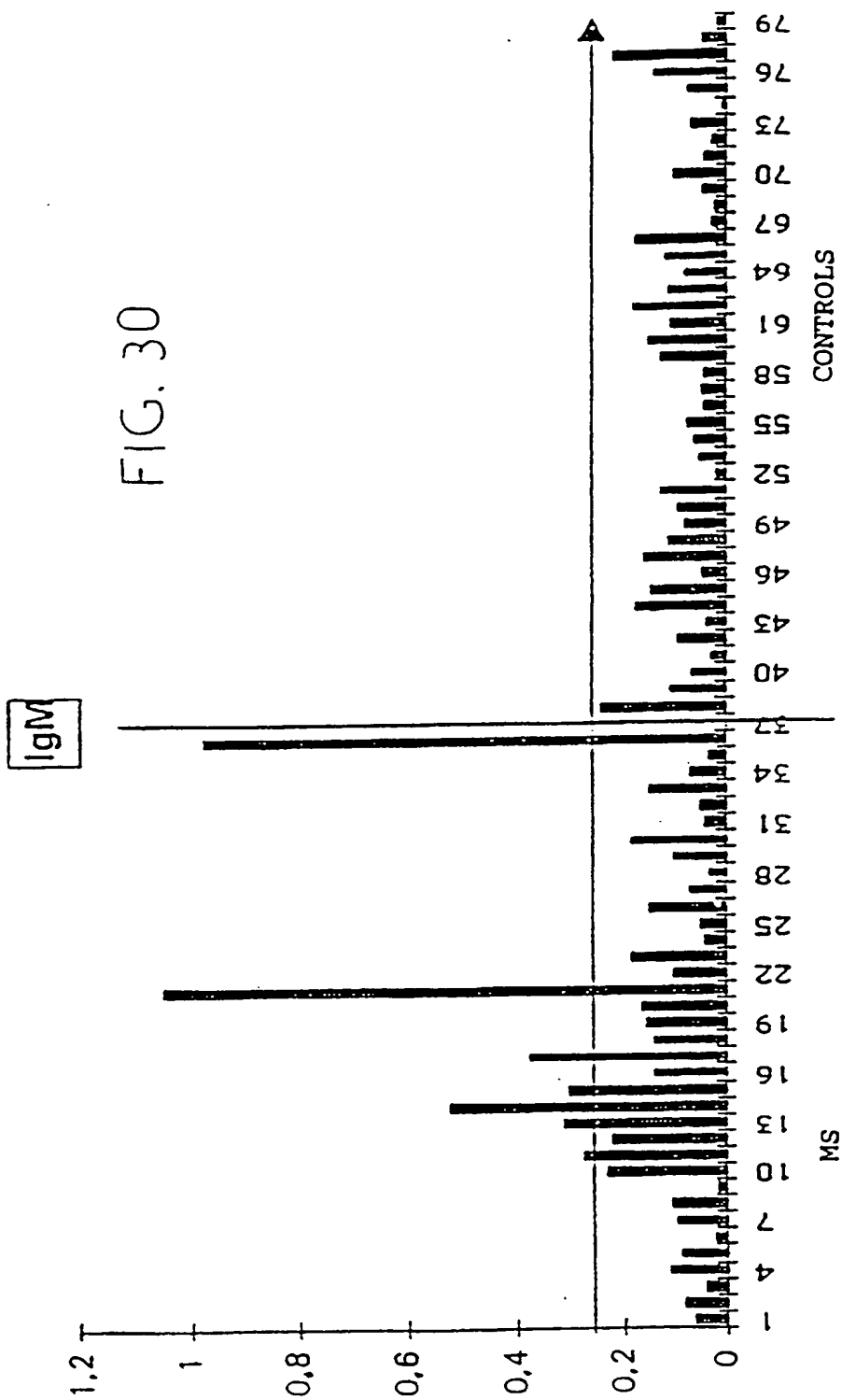
SEQ ID NO 39 (POL2B)



30/69

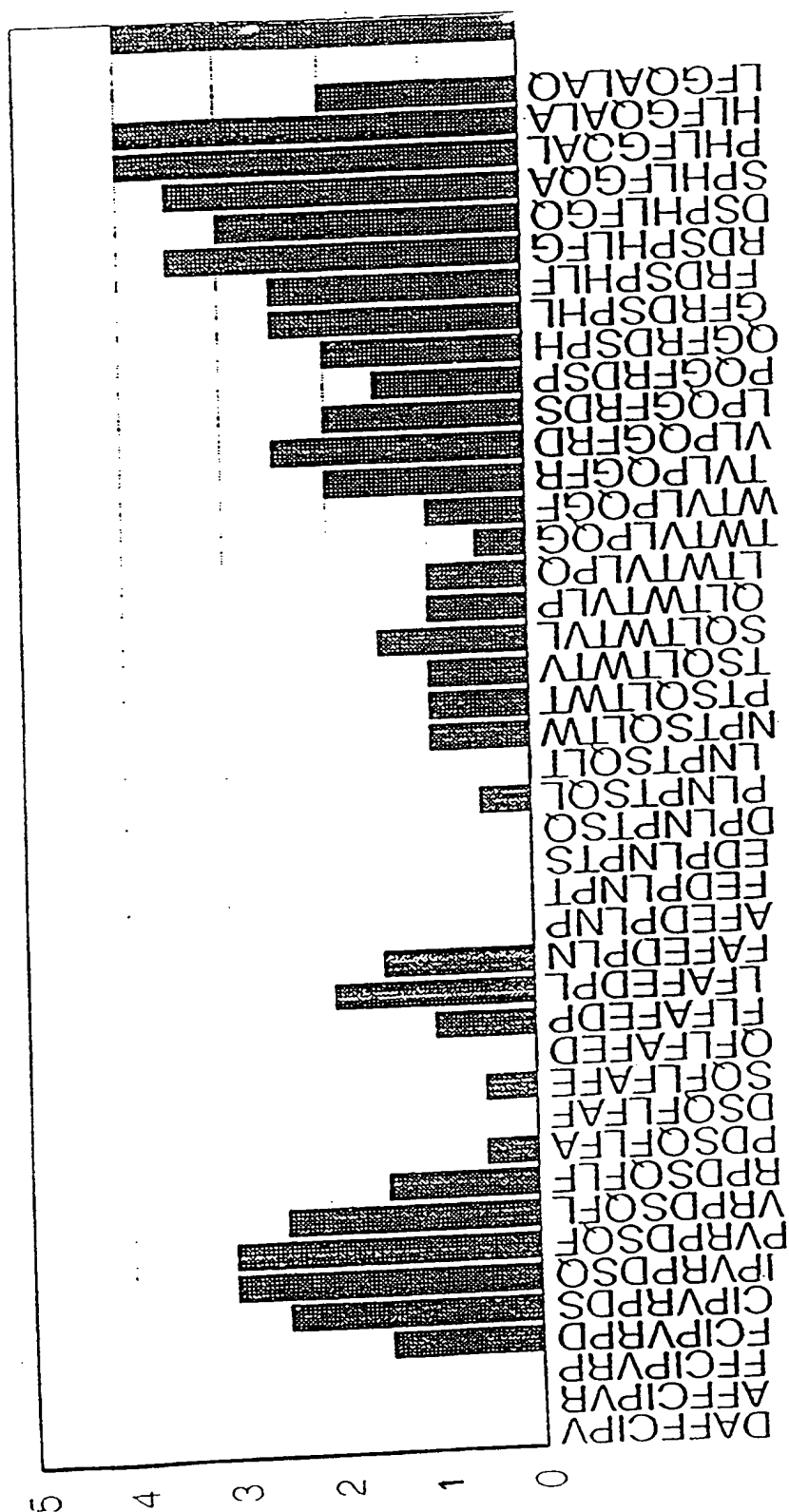


31/69



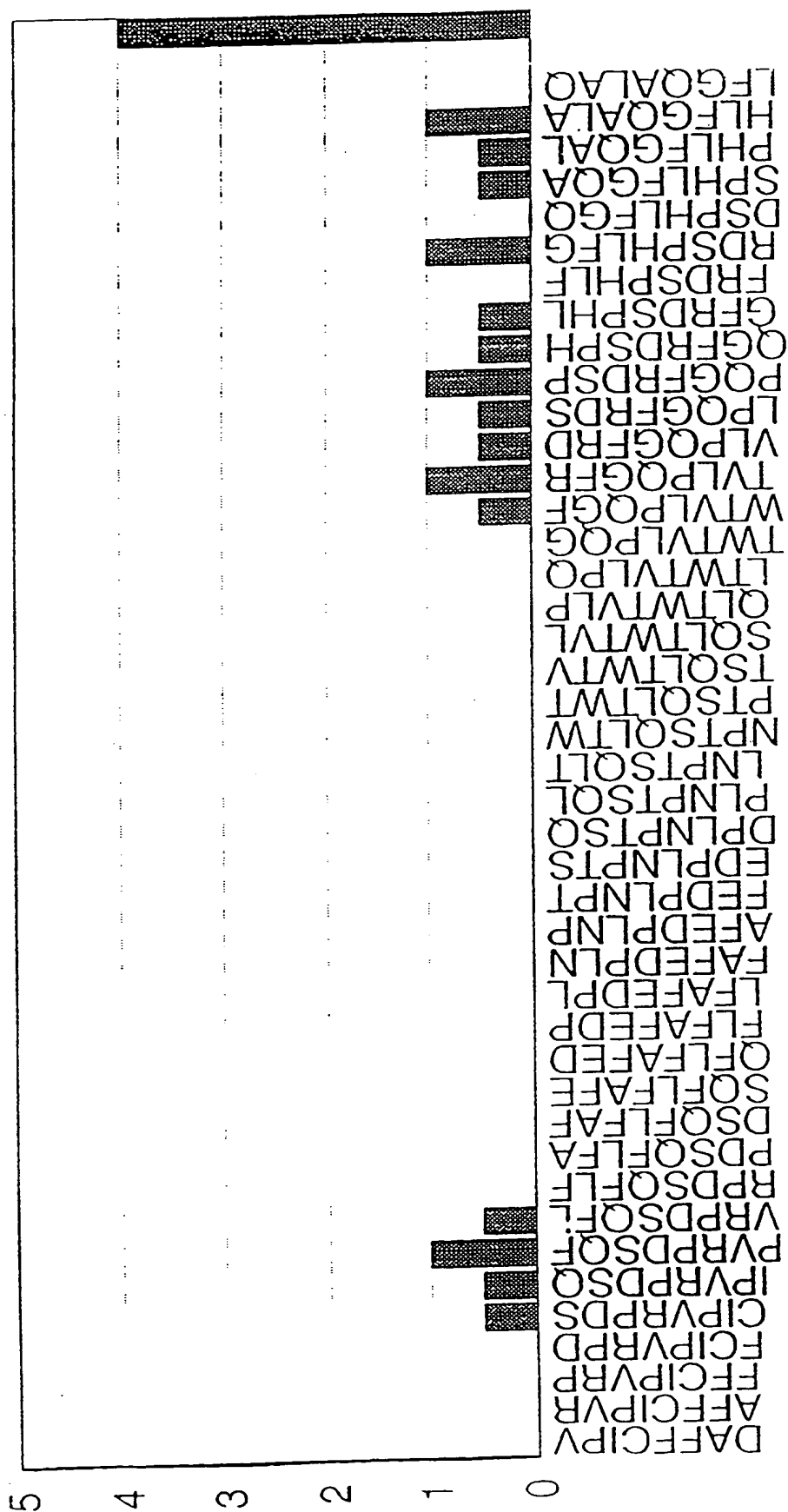
32/69

FIG. 31

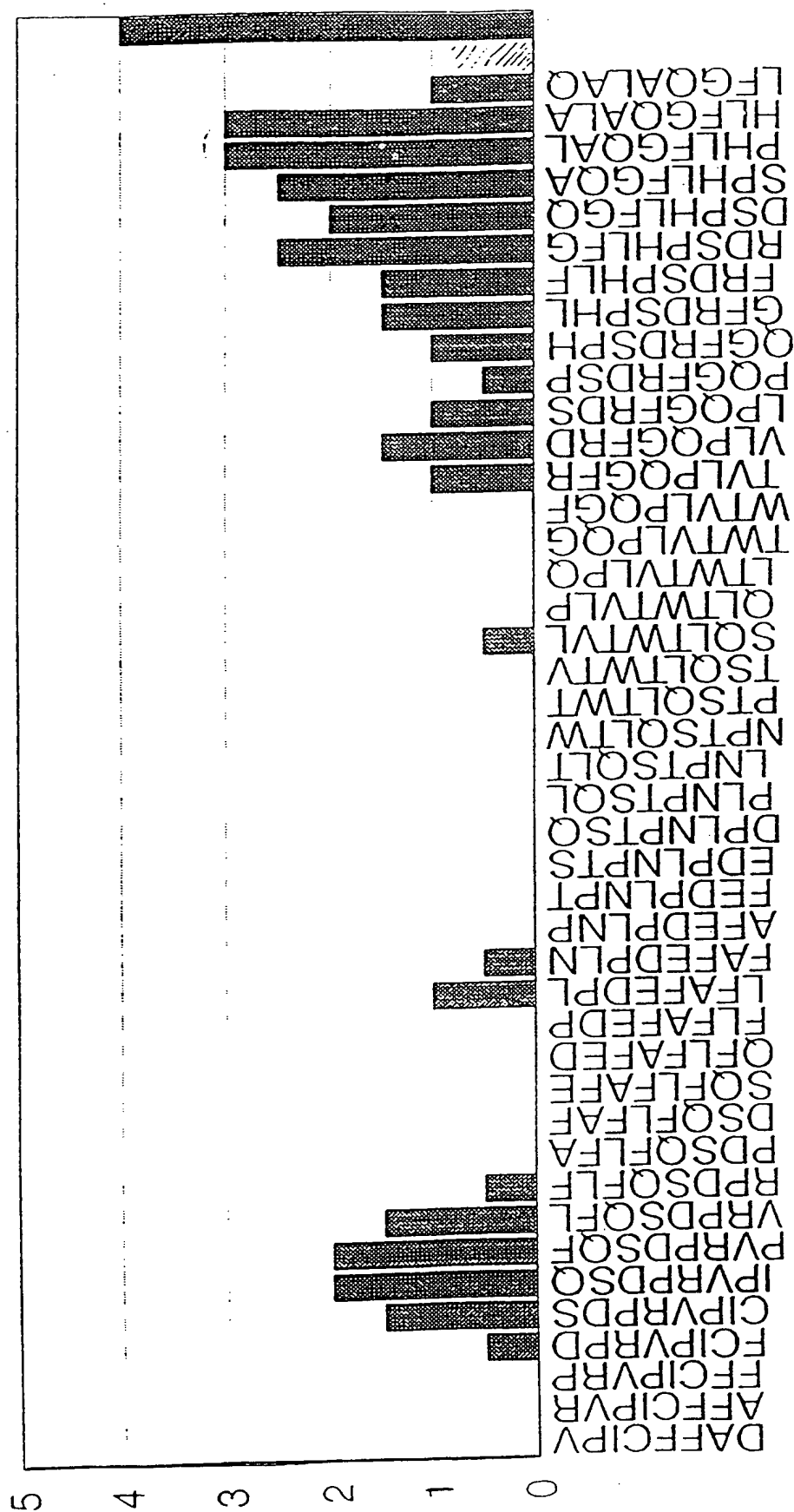
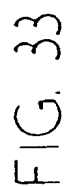


33/69

FIG. 32



34/69



*35/69*

FIG. 34

Cys-Ile-Pro-Val-Arg-Pro-Asp-Ser-Gln-Phe-Leu SEQ ID NO 41

Val-Leu-Pro-Gln-Gly-Phe-Arg-Asp-Ser-Pro-His-Leu-Phe-Gly-  
Gln-Ala-Leu-Ala SEQ ID NO 42

Leu-Phe-Ala-Phe-Glu-Asp-Pro-Leu  
Phe-Ala-Phe-Glu-Asp-Pro-Leu-Asn

SEQ ID NO 43

SEQ ID NO 44

FIG 35

36/69

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
CTTCCCCAAC	TAATAAGGAC	CCCCCTTTCA	ACCCAAACAG	TCCAAAAGGA	50
L P Q L	I R T	P L S	T Q T V	Q K D	
F P N	. . G P	P F Q	P K Q	S K R T	
S P T	N K D	P P F N	P N S	P K G	
CATAGACAAA	GGAGTAAACA	ATGAACCAAA	GAGTGCCAAT	ATTCCCTGGT	100
I D K	G V N N	E P K	S A N	I P W L	
. T K	E . T	M N Q R	V P I	F P G	
H R Q R	S K Q	. T K	E C Q Y	S L V	
TATGCACCTT	CCAAGCGGTG	GGAGAAGAAT	TCGGCCCAGC	CAGAGTGCAT	150
C T L	Q A V	G E E F	G P A	R V H	
Y A P S	K R W	E K N	S A Q P	E C M	
M H P	P S G G	R R I	R P S	Q S A C	
GTACCTTTTT	CTCTCTCACA	CITGAAGCAA	ATTAAAATAG	ACNTAGGINA	200
V P F S	L S H	L K Q	I K I D	X G X	
Y L F	L S H T	. S K	L K .	T . V N	
T F F	S L T	L E A N	. N R	X R X	
ATTINICAGAT	AGCCCTGATG	GYTATATTGA	TGTTTTACAA	GGATTAGGAC	250
X S D	S P D G	Y I D	V L Q	G L G Q	
X Q I	A L M	X I L M	F Y K	D . D	
I X R .	P . W	L Y .	C F T R	I R T	
AATCCTTTGA	TCTGACATGG	AGAGATATAA	TATTACTGCT	AAATCAGACG	300
S F D	L T W	R D I I	L L L	N Q T	
N P L I	. H G	E I .	Y Y C .	I R R	
I L .	S D M E	R Y N	I T A	K S D A	
CTAACCTCAA	ATGAGAGAAG	TGCTGCCATA	ACTGGAGCCC	GAGAGTTTGG	350
L T S N	E R S	A A I	T G A R	E F G	
. P Q	M R E V	L P .	L E P	E S L A	
N L K	. E K	C C H N	W S P	R V W	
CAATCTCTGG	TATCTCAGTC	AGGTCATGA	TAGGATGACA	ACGGAGGAAA	400
N L W	Y L S Q	V N D	R M T	T E E R	
I S G	I S V	R S M I	G . Q	R R K	
Q S L V	S Q S	G Q .	. D D N	G G K	
GAGAACGATT	CCCCACAGGG	CAGCAGGCAG	TTCCCACTGT	AGCTCCTCAT	450
E R F	P T G	Q Q A V	P S V	A P H	
E N D S	P Q G	S R Q	F P V .	L L I	
R T I	P H R A	A G S	S Q C	S S S L	
TGGGACACAG	AATCAGAACA	TGGAGATTGG	TGCGGCAGAC	ATTTA	495
W D T E	S E H	G D W	C R R H	L	
G T Q	N Q N M	E I G	A A D	I	
G H R	I R T	W R L V	P Q T	F	

FIG 36

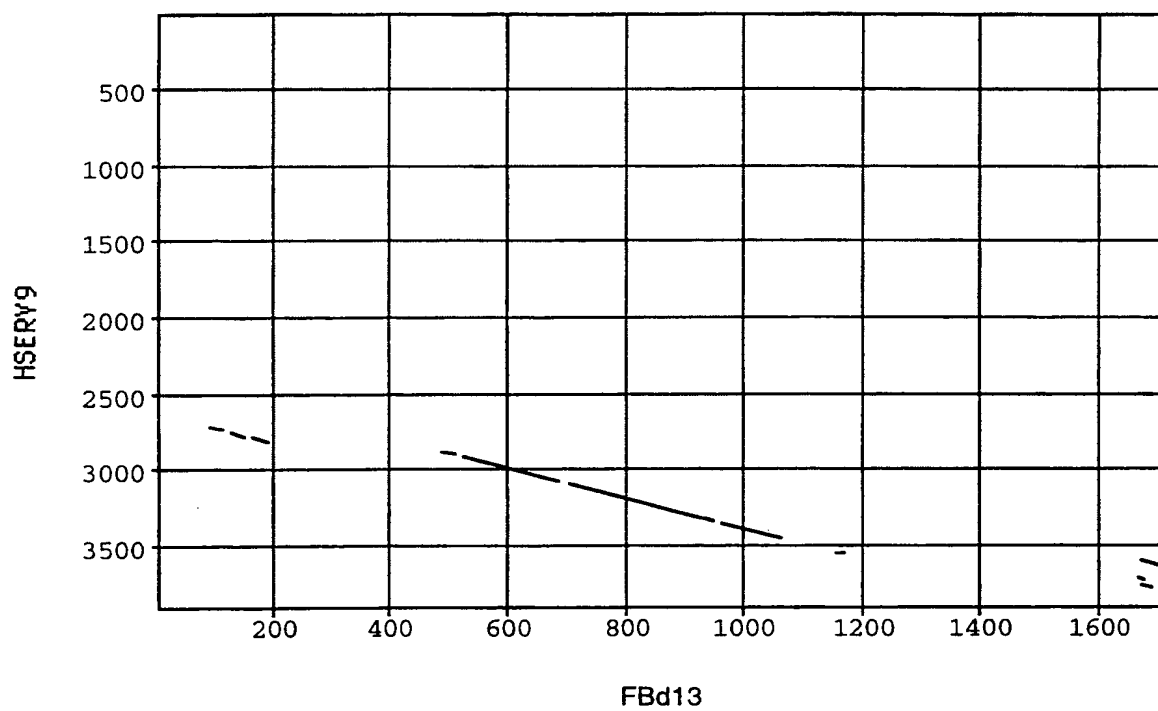
37/69

	10	20	30	40	50	
	1234567890	1234567890	1234567890	1234567890	1234567890	
	CTTCCCCAAC	TAATAAGGAC	CCCCCTTTCA	ACCCAAACAG	TCCAAAAGGA	50
	L P Q L	I R T	P L S	T Q T V	Q K D	
	CATAGACAAA	GGAGTAAACA	ATGAACCAAA	GAGTGCCAAT	ATTCCCTGGT	100
	I D K	G V N N	E P K	S A N	I P W L	
	TATGCACCOCT	OCAAGGGGTG	GGAGAAGAAT	TCGGCCCCAGC	CAGAGTGCAT	150
	C T L	Q A V	G E E F	G P A	R V H	
	GTACCTTTTT	CTCTCTCACA	CTTGAAGCAA	ATTAAATAG	ACCTAGGTAA	200
	V P F S	L S H	L K Q	I K I D	L G K	
	ATTCTCAGAT	AGCCCTGATG	GYTATATTGA	TGTTTTACAA	GGATTAGGAC	250
	F S D	S P D G	Y I D	V L Q	G L G Q	
	AATCCTTTGA	TCTGACATGG	AGAGATATAA	TATTACTGCT	AAATCAGACG	300
	S F D	L T W	R D I I	L L L	N Q T	
	CTAACCTCAA	ATGAGAGAAG	TGCTGCCATA	ACTGGAGCCC	GAGAGTTTGG	350
	L T S N	E R S	A A I	T G A R	E F G	
	CAATCTCTGG	TATCTCAGTC	AGGTCATGA	TAGGATGACA	ACGGAGGAAA	400
	N L W	Y L S Q	V N D	R M T	T E E R	
	GAGAACGATT	CCCCACAGGG	CAGCAGGCAG	TTCOCAGTGT	AGCTOCTCAT	450
	E R F	P T G	Q Q A V	P S V	A P H	
	TGGGACACAG	AATCAGAACA	TGGAGATTGG	TGCOGCAGAC	ATTTACAAC T	500
	W D T E	S E H	G D W	C R R H	L Q L	
	TGCGTGCTAN	AAGGACTINAG	GAAAAC TAGG	AAGACTPANGA	ATTATTCAAN	550
	A C X	K D X G	K L G	R L X	I I Q X	
	GATGTCCACT	ANNACACAGG	GGAAAGGAAG	AAAATCTTAC	TGCOCTTTCTG	600
	C P L	X H R	G K E E	N P T	A F L	
	GAGAGACTAA	GGGAGGCATT	GAGGAAGCAT	ACCAGGCAAG	TGGACATTGG	650
	E R L R	E A L	R K H	T R Q V	D I G	
	AGGCTCTGGA	AAAGGGAAAA	GTIGGGCAAA	TTATATGCOCT	AATAGGGCTT	700
	G S G	K G K S	W A N	Y M P	N R A C	
	GCTTCCAGTG	CAGTCTACAA	GAAGCCTTTA	GAAAAGATTG	TOCAAGTAGA	750
	F Q C	S L Q	G R F R	K D C	P S R	
	AATAAGCCGC	CCCTCGTCCA	TGCCCCCTTAT	GTCAAGGGAA	TCACTGGAAG	800
	N K P P	L V H	A P Y	V K G I	T G R	
	GCCTACTGCC	CCAGGGGACG	AAGGTCTCT	GAGTCAGAAG	CCACTTAACT	850
	P T A	P G D E	G P L	S Q K	P L T	
GA						852



38/69

FIG 37



39/69

FIG 38  
a

	10	20	30	40	50	
	1234567890	1234567890	1234567890	1234567890	1234567890	
	AAGGAACTC	AGAAAGCCAA	TACCCATTTA	GTAAGATGGA	CACCAGAAGC	50
	K E T Q	K A N	T H L	V R W T	P E A	
	R K L	R K P I	P I .	. D G	H Q K Q	
	G N S	E S Q	Y P F S	K M D	T R S	
	AGAAGCAGCT	TTCCAGGCC	TAAAGAAATC	CCTAACCCAA	GCCCCAGTGT	100
	E A A	F Q A L	K K S	L T Q	A P V L	
	K Q L	S R P	. R N P	. P K	P Q C	
	R S S F	P G P	K E I	P N P S	P S V	
	TAAGCTTGCC	AACGGGGCAA	GACTTTTCTT	TATATGTCAC	AGAAAAACAG	150
	S L P	T G Q	D F S L	Y V T	E K Q	
	. A C Q	R G K	T F L	Y M S Q	K N R	
	K L A	N G A R	L F F	I C H	R K T G	
	GAATAGCTCT	AGGAGTCCTT	ACACAGGTC	AAGGGACAAG	CTTGCAACCT	200
	E . L .	E S L	H R S	K G Q A	C N L	
	N S S	R S P Y	T G P	R D K	L A T C	
	I A L	G V L	T Q V Q	G T S	L Q P	
	GTGGCATAAC	TGAGTAAGGA	AACTGATGTA	NIGGCAAAGG	GTTGGCCTCA	250
	W H T	. V R K	L M X	W Q R	V G L I	
	G I P	E . G N	. C X	G K G	L A S	
	V A Y L	S K E	T D V	X A K G	W P H	
	TTGTTTACAG	GTAGGGCAGC	AGTAGCAGTC	TTAGTTTCTG	AAACAGTTAA	300
	V Y R	. G S	S S S L	S F .	N S .	
	L F T G	R A A	V A V	L V S E	T V K	
	C L Q	V G Q Q	. Q S	. F L	K Q L K	
	AATAATACAG	GGAAGAGATC	TTACTGTGTG	GACATCTCAT	GATGIGAACG	350
	N N T G	K R S	Y C V	D I S .	C E R	
	I I Q	G R D L	T V W	T S H	D V N G	
	. Y R	E E I	L L C G	H L M M	. T	
	GCATACTCAC	TGCTAAAGAG	GACTTGTGGC	TGTCAGACAA	CCATTTACTT	400
	H T H C	. R G	L V A	V R Q	P F T	
	I L T	A K E	D L W L	S D N	H L L	
	A Y S L	L K R	T C G	C Q T T	I Y L	
	AAATAGCAGG	TTCTATTACT	TGAAGTGCCA	GTCCTGGGAC	TGCACATTTG	450
	I A G	S I T	. S A S	A A T	A H L	
	K . Q V	L L L	E V P	V L R L	H I C	
	N S R	F Y Y L	K C Q	C C D	C T F V	
	TGCAACTCTT	AACCCAGCCA	CATTTCCTCC	AGACAATGAA	GAAAAGATAG	500
	C N S .	P S H	I S S	R Q .	R K D R	
	A T L	N P A T	F L P	D N E	E K I E	
	Q L L	T Q P	H F F Q	T M K	K R .	

40/69

FIG38  
b

	10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	
AACATAACTG	TCAACAAGTA	ATTGCTCAAA	CCTATGCTGC	TCGAGGGGAC		550
T . L S T S N	C S N L C C	S R G P				
H N C Q Q V	I A Q T Y A A	R G D				
N I T V N K .	L L K P M L L	E G T				
CTTCTAGAGG	TTCCTTGAC	TGATCCCGAC	CTCAACTTGT	ATACTGATGG		600
S R G S L D .	S R P Q L V Y .	W				
L L E V P L T	D P D L N L Y	T D G				
F . R F P . L	I P T S T C	I L M E				
AAGTTCCTTG	GCAGAAAAAG	GACTTTGAAA	AGCGGGGTAT	GCAGTGATCA		650
K F L G R K R	T L K S G V C	S D Q				
S S L A E K G	L . K A G Y	A V I S				
V P W Q K K	D F E K R G M	Q . S				
GIGATAATGG	AATACTTGAA	AGTAATCGCC	TCACTCCAGG	AACTAGTGCT		700
. . W N T . K .	S P H S R N .	C S				
D N G I L E	S N R L T P G	T S A				
V I M E Y L K	V I A S L Q E	L V L				
CACCTGGCAG	AACTAATAGC	CCTCACTTGG	GCACTAGAAT	TAGGAGAAGG		750
P G R T N S	P H L G T R I	R R R				
H L A E L I A	L T W A L E L	G E G				
T W Q N . .	P S L G H .	N . E K E				
AAAAAGGGTA	AATATATATT	CAGACTCTAA	GTATGCTTAC	CTAGTCTTCC		800
K K G K . Y I F	R L . V C L P	S P P				
K R V N I Y S	D S K Y A Y	L V L H				
K G . I Y I	Q T L S M L T	. S S				
ATGCCCATGC	AGCAATATGG	AGAGAGAGGG	AATTCCTAAC	TTCTGAGGGA		850
C P C S N M E	R E G I P N	F . G N				
A H A A I W	R E R E F L T	S E G				
M P M Q Q Y G	E R G N S .	L L R E				
ACACCTATCA	ACCATCAGGG	AAGCCATTAG	GAGATTATTA	TTGGCTGTAC		900
T Y Q P S G	K P L G D Y Y	W L Y				
T P I N H Q G	S H . E I I I	G C T				
H L S T I R E	A I R R L L	L A V Q				
AGAAACCTAA	AGAGGTGGCA	GTCTTACACT	GCCAGGGTCA	TCAGGAAGAA		950
R N L K R W Q	S Y T A R V I	R K K				
E T . R G G S	L T L P G S	S G R R				
K P K E V A	V L H C Q G H	Q E E				
GAGGAAAGGG	AAATAGAAGG	CAATCGCCAA	GCGGATATTG	AAGCAAAAAA		1000
R K G K . K A	I A K R I L	K Q K K				
G K G N R R	Q S P S G Y .	S K K				
E E R E I E G	N R Q A D I E	A K K				

41/69

FIG 38  
C

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
AGCCGCAAGG	CAGGACTCTC	CATTAGAAAT	GCCTATAGAA	GGACCCCTAG	1050
P Q G R T L H . K C L . K D P .					
S R K A G L S I R N A Y R R T P S					
A A R Q D S P L E M L I E G P L V					
TATGGGGTAA	TCCCCTCTGG	GAAACCAAGC	CCAGTACTC	AGCAGGAAAA	1100
Y G V I P S G K P S P S T Q Q E K					
M G . S P L G N Q A P V L S R K N					
W G N P L W E T K P Q Y S A G K					
ATAGAATAGG	AAACCTCACA	AGGACATACT	TTCCTCCCCT	CCAGATGGCT	1150
. N R K P H K D I L S S P P D G .					
R I G N L T R T Y F P P L Q M A					
I E . E T S Q G H T F L P S R W L					
AGCCACTGAG	GAAGGAA				1167
P L R K E					
S H . G R					
A T E E G					

42/69

FIG 39

a

	10	20	30	40	50	
	1234567890	1234567890	1234567890	1234567890	1234567890	
	A A C T T G C G T G	C T A G A A G G A C	T A A G G A A A A C	T A G G A A G A C T	A T G A A T T A T T	50
	N L R A	R R T	K E N	. E D Y	E L F	
	T C V	L E G L	R K T	R K T	M N Y S	
	L A C	. K D	. G K L	G R L	. I I	
	C A A T G A T G T C	C A C T A T A A C A	C A G G G A A A G	G A A G A A A A T C	C T A C T G C C T T	100
	N D V	H Y N T	G E R	K K I	L L P F	
	M M S	T I T	Q G K G	R K S	Y C L	
	Q . C P L	. H R G K	E E N P	T A F		
	T C T G C A G A C A	C T A A G G G A G G	C A T T G A G G A A	G C A T A C C A G G	C A A G T G G A C A	150
	W R D	. G R H	. G S	I P G	K W T	
	S G E T	K G G	I E E	A Y Q A	S G H	
	L E R	L R E A	L R K	H T R	Q V D I	
	T T G G A G G C T C	T G G A A A G G G	A A A G T T G G G	C A A A T T G A A T	G C C T A A T A G G	200
	L E A L	E K G	K V G	Q I E C	L I G	
	W R L	W K R E	K L G	K L N A	. . G	
	G G S	G K G	K S W A	N . M	P N R	
	G C T T G C T T C C	A G T G C A G T C T	A C A A G G A C G C	T T T A G A A A A G	A T T G T C C A A G	250
	L A S	S A V Y	K D A	L E K	I V Q V	
	L L P	V Q S	T R T L	. K R	L S K	
	A C F Q	C S L	Q G R	F R K D	C P S	
	T A G A A A T A A G	C C G C C C C T C G	T C C A T G C C C C	T T A T G T C A A G	G G A A T C A C T G	300
	E I S	R P S	S M P L	M S R	E S L	
	. K .	A A P R	P C P	L C Q G	N H W	
	R N K	P P L V	H A P	Y V K	G I T G	
	G A A G G C C T A C	T G C C C C A G G G	G A C G A A G G T C	C T C T G A G T C A	G A A G C C A C T A	350
	E G L L	P Q G	T K V	L . V R	S H .	
	K A Y	C P R G	R R S	S E S	E A T N	
	R P T	A P G	D E G P	L S Q	K P L	
	A C C T G A T G A T	C C A G C A G C A G	G A C T G A G G G T	G C C C G G G G C A	A G T G C C A G C C	400
	P D D	P A A G	L R V	P G A	S A S P	
	L M I	Q Q Q	D . G C	P G Q	V P A	
	T . .	S S S R	T E G	A R G K	C Q P	
	C A T G C C A T C A	C C C T C A G A G C	C C C G G G T A T G	T T T G A C C A T T	G A G A G C C A G G	450
	C H H	P Q S	P G Y V	. P L	R A R	
	H A I T	L R A	P G M	F D H .	E P G	
	M P S	P S E P	R V C	L T I	E S Q E	
	A A G T T A A C T G	T C T C C T G G A C	A C T G G G C C A G	C C T T C T C A G T	C T T A C T T T C C	500
	K L T V	S W T	L A Q	P S Q S	Y F P	
	S .	L S P G H	W R S	L L S	L T F L	
	V N C	L L D	T G A A	F S V	L L S	

43/59

FIG 39  
b

	10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	
TGTCCTCAGAC	AATTGTCTCTC	CAGATCTGTG	ACTATCCGAG	GGGTCTTAAG		550
V P D N C P P	D L S L S E G S	.	D			
S Q T I V L	Q I C H Y P R	G P K				
C P R Q L S S	R S V T I R G	V L R				
ACAGCCAGTC	ACTACATACT	TCTCTCAGCC	ACTAAGTTGT	GACTGGGGAA		600
S Q S L H T	S L S H . V V	T G E				
T A S H Y I L	L S A T K L . L G N					
Q P V T T Y F	S Q P L S C D W G T					
CTTTACTCTT	TTACATGCT	TTTCTAATTA	TGCTGAAAG	CCCCACTCCC		650
L Y S F H M L	F . L C L K A	P L P				
F T L F T C F	S N Y A . K P H S L					
L L F S H A	F L I M P E S	P T P				
TTGTTAGGGA	GAGACATTTT	AGCAAAAGCA	GGGGCCATTA	TACACCTGAA		700
C . G E T F .	Q K Q G P L Y T . T					
V R E R H F	S K S R G H Y T P E					
L L G R D I L	A K A G A I I H L N					
CATAGGAAAA	GGAATACCCA	TTTGCTGTCC	CCTGCTTGAG	GAAGGAATTA		750
. E K E Y P	F A V P C L R K E L					
H R K R N T H	L L S P A . G R N .					
I G K G I P I	C C P L L E E G I N					
ATCCTGAAGT	CTGGGCAATA	GAAGACAAT	ATGGACAAGC	AAAGAATGCC		800
I L K S G Q .	K D N M D K Q R M P					
S . S L G N R	R T I W T S K E C P					
P E V W A I	E G Q Y G Q A K N A					
CGTCTGTTC	AAGTTAACT	AAAGGATTTCT	GCCTCCTTTT	CCTACCAAAG		850
V L F K L N .	R I L P P F P T K G					
S C S S . T	K G F C L L S L P K					
R P V Q V K L	K D S A S F P Y Q R					
GAAGTACCTT	CTTAGACCCG	AGGCCCTACA	AGGACTCAAA	AGATTGTTAA		900
S T L L D P	R P Y K D S K D C .					
E V P S . T R	G P T R T Q K I V K					
K Y P L R P E	A L Q G L K R L L R					
GGACCTAAAA	GOCCAAGGCC	TAGTAAACC	ATGCAGTAGC	CCCTGCAATA		950
G P K S P R P	S K T M Q . P L Q Y					
D L K A Q G L	V K P C S S P C N T					
T . K P K A . .	N H A V A P A I					
CTCCAATTTT	AGGAGTAAGG	AAACCAACG	GACAGTGGAG	GTTAGTGCAA		1000
S N F R S K E	T Q R T V E V S A R					
P I L G V R	K P N G Q W R L V Q					
L Q F . E . G	N P T D S G G . C K					

FIG 39  
C

44/69

	10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	
GATCTCAGGA	TTATTAAATGA	GGCTGTTTTT	CCTCTATAACC	CAGCTGTATC		1050
S Q D Y . .	G C F S	S I P	S C I			
D L R I	I N E	A V F	P L Y P	A V S		
I S G	L L M R	L F F	L Y T	Q L Y L		
TAGCCCTTAT	ACTCTGCTTT	CCCTAATAACC	AGAGGAAGCA	CAGTAGTTTA		1100
. P L Y	S A F	P N T	R G S R	V V Y		
S P Y	T L L S	L I P	E E A	E . F T		
A L I	L C F	P . Y Q	R K Q	S S L		
CAGTCTGGA	CCTTAAGGAT	GGCTCTTTCT	GCATCCCTGT	ACATCCTGAT		1150
S P G	P . G C	L F L	H P C	T S . F		
V L D	L K D	A S F C	I P V	H P D		
Q S W T	L R M	P L S	A S L Y	I L I		
TCTCAATTCT	TGTTTGCTT	TGAAGATCCT	TTGAACCCAA	TGCTCTCAATT		1200
S I L	V C L	. R S F	E P N	V S I		
S Q F L	F V F	E D P	L N P M	S Q F		
L N S	C L S L	K I L	. T Q	C L N S		
CACCTGGACT	GTTTTACCCC	AGGGGTTCCG	GGATAGCCCC	CATCTATTTG		1250
H L D C	F T P	G V P	G . P P	S I W		
T W T	V L P Q	G F R	D S P	H L F G		
P G L	F Y P	R G S G	I A P	I Y L		
GGCAGGCATT	AGCCCAAGAC	TTGAGCCAAT	TCTCATACCT	GGACATCTTG		1300
P G I	S P R L	E P I	L I P	G H L V		
Q A L	A Q D	L S Q F	S Y L	D I L		
A R H .	P K T	. A N	S H T W	T S C		
TCCTTGGTA	TGGGATGATT	TAATTTTAGC	CACCCGTTCA	GAAACCTTGT		1350
L R Y	G M I	. F . P	P V Q	K P C		
S F G M	G . F	N F S	H P F R	N L V		
P S V	W D D L	I L A	T R S	E T L C		
GCCATCAAGC	CACCCAAGCG	TTCTTAAATT	TCCTCACTCC	GTGTGGCTAC		1400
A I K P	P K R	S . I	S S L R	V A T		
P S S	H P S V	L K F	P H S	V W L Q		
H Q A	T Q A	F L N F	L T P	C G Y		
AAGGTTTCCA	AACCAAAGGC	TCAGCTCTGC	TCACAGCAGG	TTAAATACTT		1450
R F P	N Q R L	S S A	H S R	L N T .		
G F Q	T K G	S A L L	T A G	. I L		
K V S K	P K A	Q L C	S Q Q V	K Y L		
AGGGTTAAAA	TTATCCAAAG	GCACCAGGGC	CCTCTGTGAG	GAATGTATCC		1500
G . N	Y P K	A P G P	S V R	N V S		
R V K I	I Q R	H Q G	P L . G	M Y P		
G L K	L S K G	T R A	L C E	E C I Q		

45/69

FIG 39  
d

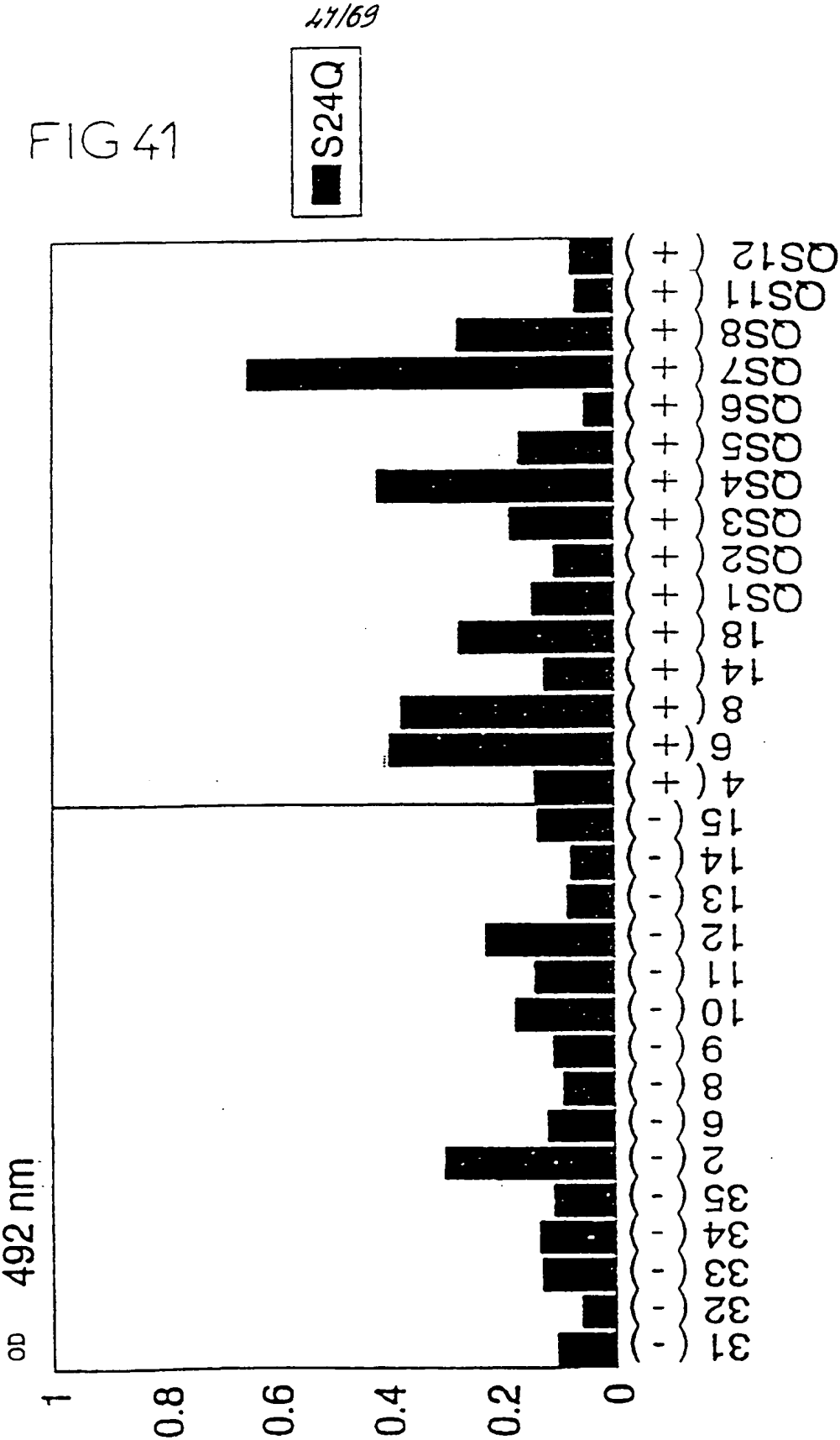
10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
AACCTGTACT	GGCTTATCIT	CATCCCAAAA	COCTAAAGCA	ACTAAGAAGG	1550
N L Y W	L I F	I P K	P . S N	. E G	
T C T	G L S S	S Q N	P K A	T K K V	
P V L	A Y L	H P K T	L K Q	L R R	
TCCTTGGCAT AACAGGTTTC TGCCGAA					1577
P W H	N R F L	P			
L G I	T G F	C R			
S L A .	Q V S	A E			



46/69

FIG 40

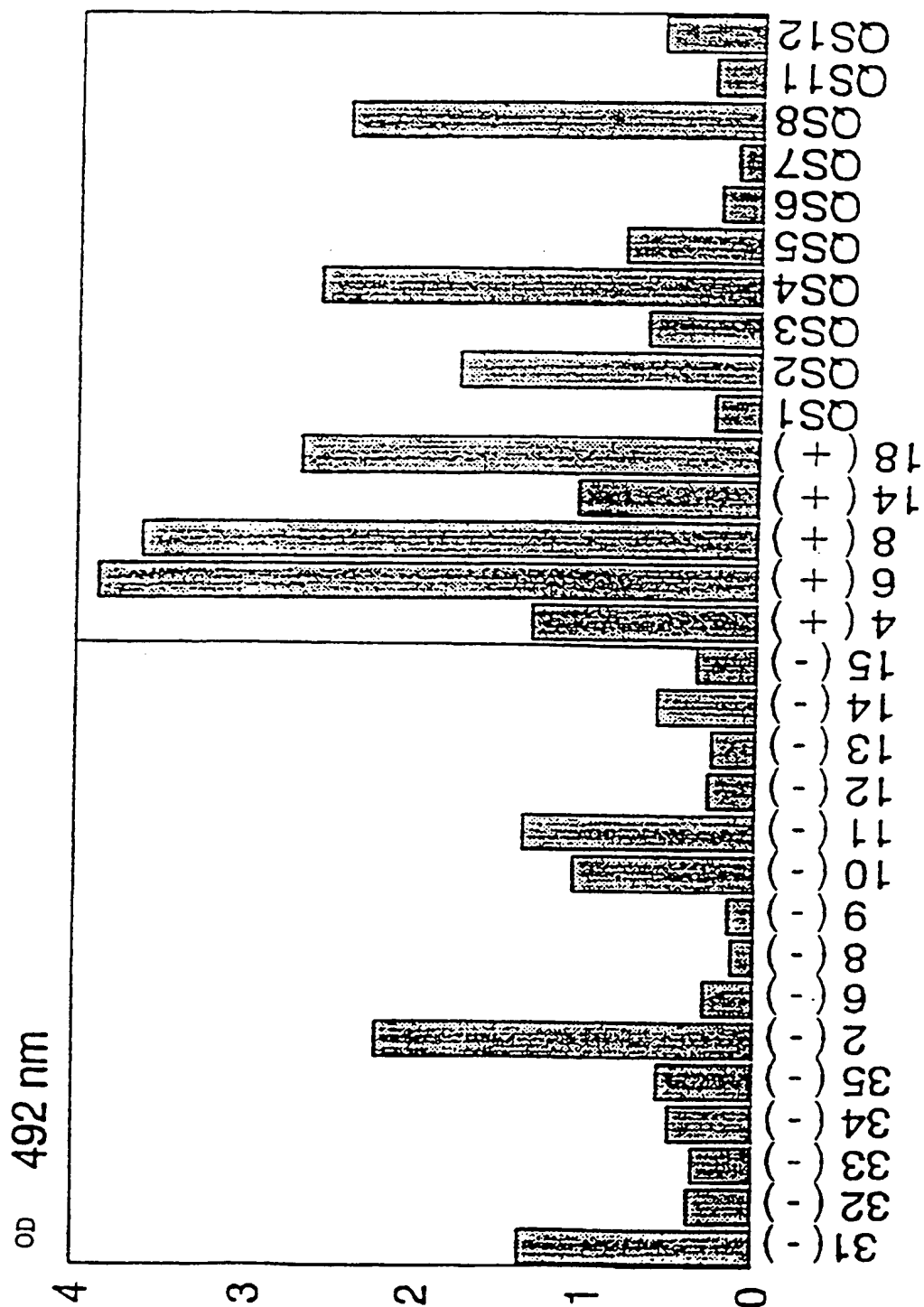
10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TCCAGCAGCA	GGACTGAGGG	TGCCCCGGGGC	AAGTGCCAGC	CCATGCCATC	50
S S S R	T E G	A R G	K C Q P	M P S	
ACCCCTCAGAG	CCCCGGGTAT	GTTTGACCAT	TGAGAGCCAG	GAAGTTAACT	100
P S E	P R V C	L T I	E S Q	E V N C	
GTCCTCTGGA	CACTGGCGCA	GCCTTCTCAG	TCTTACTTTC	CTGTCCAGA	150
L L D	T G A	A F S V	L L S	C P R	
CAATTGTCTT	CCAGATCTGT	CACTATCCGA	GGGGTCCCTAA	GACAGCCAGT	200
Q L S S	R S V	T I R	G V L R	Q P V	
CACTACATAC	TTCTCTCAGC	CACTAAGTTG	TGACTGGGGA	ACTTTACTCT	250
T T Y	F S Q P	L S C	D W G	T L L F	
TTTCACATGC	TTTTCTAAIT	ATGCGTGAAA	GCCCCACTCC	CTTGTTAGGG	300
S H A	F L I	M P E S	P T P	L L G	
AGAGACATTT	TAGCAAAAGC	AGGGGGCCATT	ATACACCTGA	ACATAGGAAA	350
R D I L	A K A	G A I	I H L N	I G K	
AGGAATACCC	ATTTGCTGTC	CCCTGCTTGA	GGAAGGAATT	AATCCTGAAG	400
G I P	I C C P	L L E	E G I	N P E V	
TCTGGGCAAT	AGAAGGACAA	TATGGACAAG	CAAAGAATGC	CCGTCTGT	450
W A I	E G Q	Y G Q A	K N A	R P V	
CAAGTTAAAC	TAAAGGATTC	TGCCTCCTTT	CCCTACCAAA	GGAAGTACCC	500
Q V K L	K D S	A S F	P Y Q R	K Y P	
TCTTAGACCC	GAGGCCCTAC	AAGGACTCAA	AAGATTGTTA	AGGACCT	547
L R P	E A L Q	G L K	R L L	R T	



48/69

FIG 42

S24Q



49/69

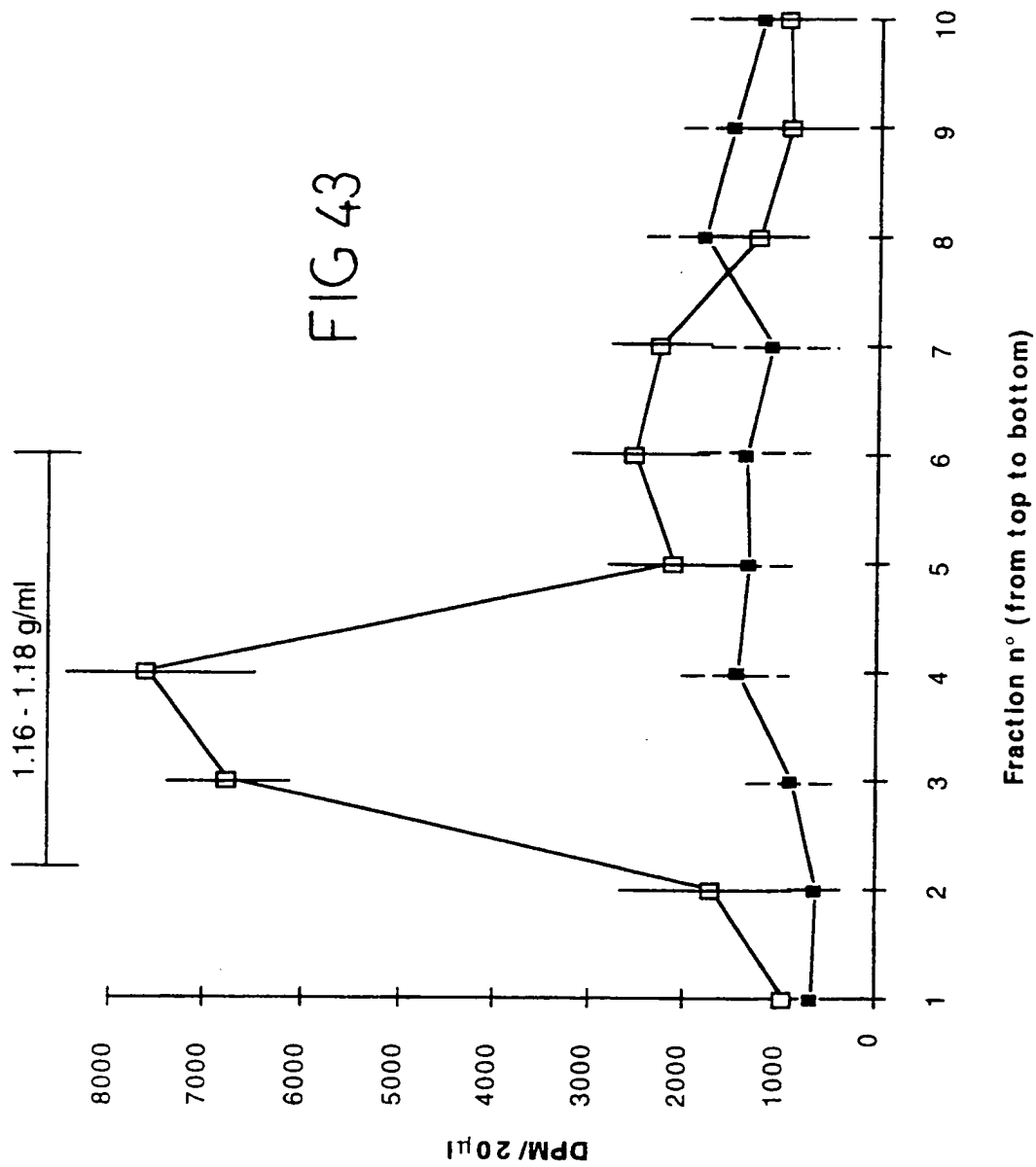


FIG 44

PAN-UO	5'-TGGAAAGTGTACCC-3' C G A	gttaaaatagtcacacccctgttcgaaatgcagctggcccatatctctgcagccattcggcaagctttcccccaatgcactattcttcag	3'-atgtacctactgtacgac-5' C G G	tacatggatgacattctc
PAN-UI	5'-LagTgttaccacagg-3' C G A G	W K M E R T L E E M Q L A H I L Q P I R Q A P P Q C T I L Q		X M D D I L
PAN-DI		tacaatgtgttccacagg		tacatggatgattgtat
HIV-1		Y N Y L E Q Q	W K Q E R A I E Q S S M T K I L E P P K K Q M P D I V I Y Q	X M D D L Y
MOHLV		tggaccagactccacagg	tttcaaaacagtcacacccctgtttgtatgaggccctgcacagaccttagcagacttccggatccagcaccagacttgatcctgtctacag	tacgtggatgacttactg
MPHV		W T R L E Q Q	P K M S E R T L E D E A L H R D L A D P R I Q H P D L I L L Q	X V D D L L
BRV9		tggaaagggtttccacagg	tatggccacagtccttactcttattgtcaaaatattgtgcccacagccatataaaggttagacatgccttggaacaaatgtatatatacat	tacatggatgacatccta
MSRV-cp1		W K Y L E Q Q	M A N S E R T L C Q K Y V A T A I H K V R H A M K Q M Y I I H	X M D D I L
DpV1		tggatggcttggcccaagg	gtttagggtatgcccctcatctgtttgttcaggcccttagccaaagatctaggccacttctcagtcaggc-----acctcgggtcttcaa	tatgtggatgatttactt
CpV1B		W M Y L E Q Q	P R D S E R H L E Q Q A L A K D L G H P S S P Q - - T L V L Q	X V D D L L
			gttcagggtatgcccccatctatttggccaggcatttagcccaagacttgagccaattctcatacctggac-----actcttgccttcag	
			P R D S E R H L E Q Q A L A Q D L S Q P S Y L D - - T L V L Q	
			K Y R W	
			Peron-5'-catcttittggicaggcattagc-3'	
			5'-cttgagccagttctcatacctgga-3'	

50/69

51/69

FIG 45

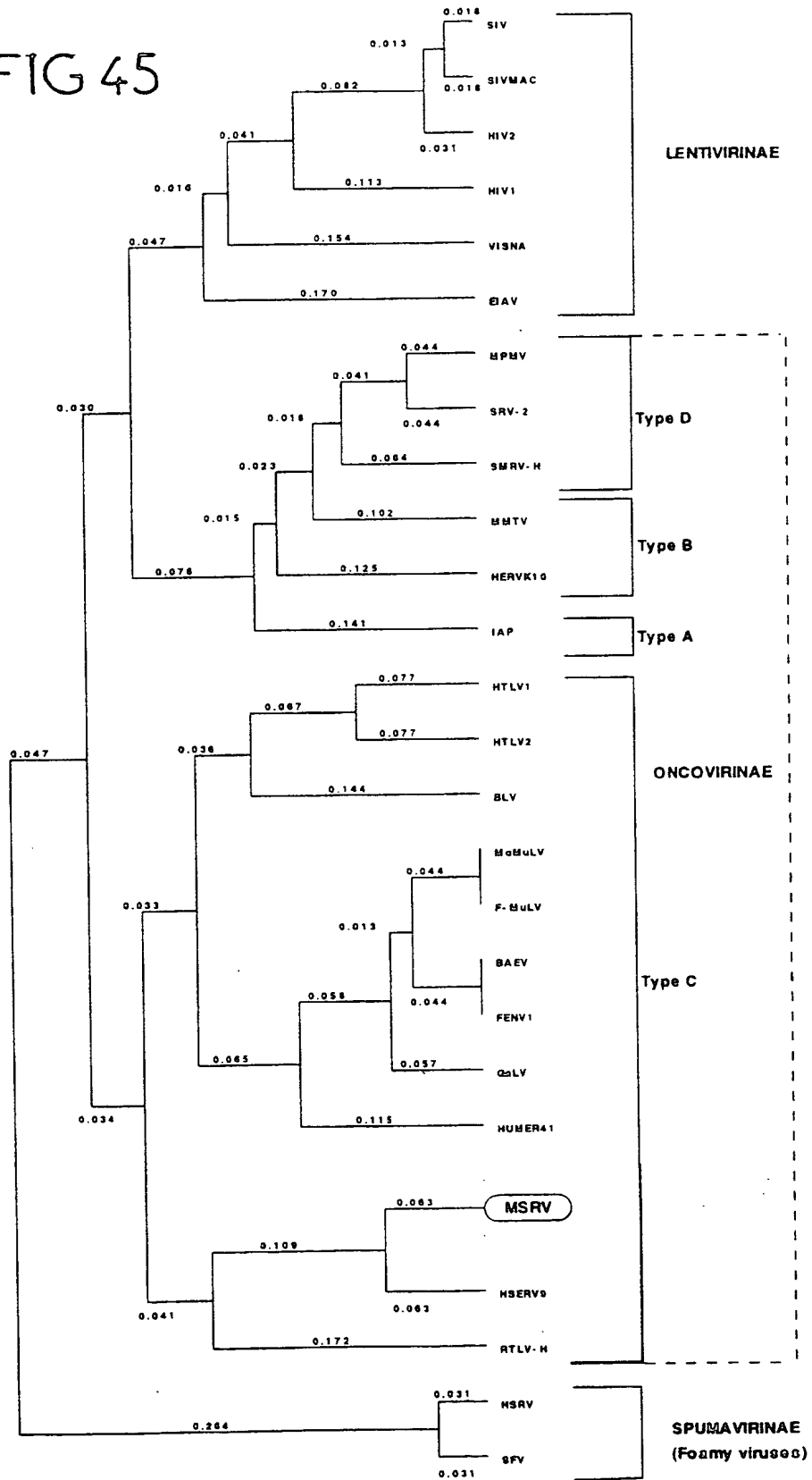


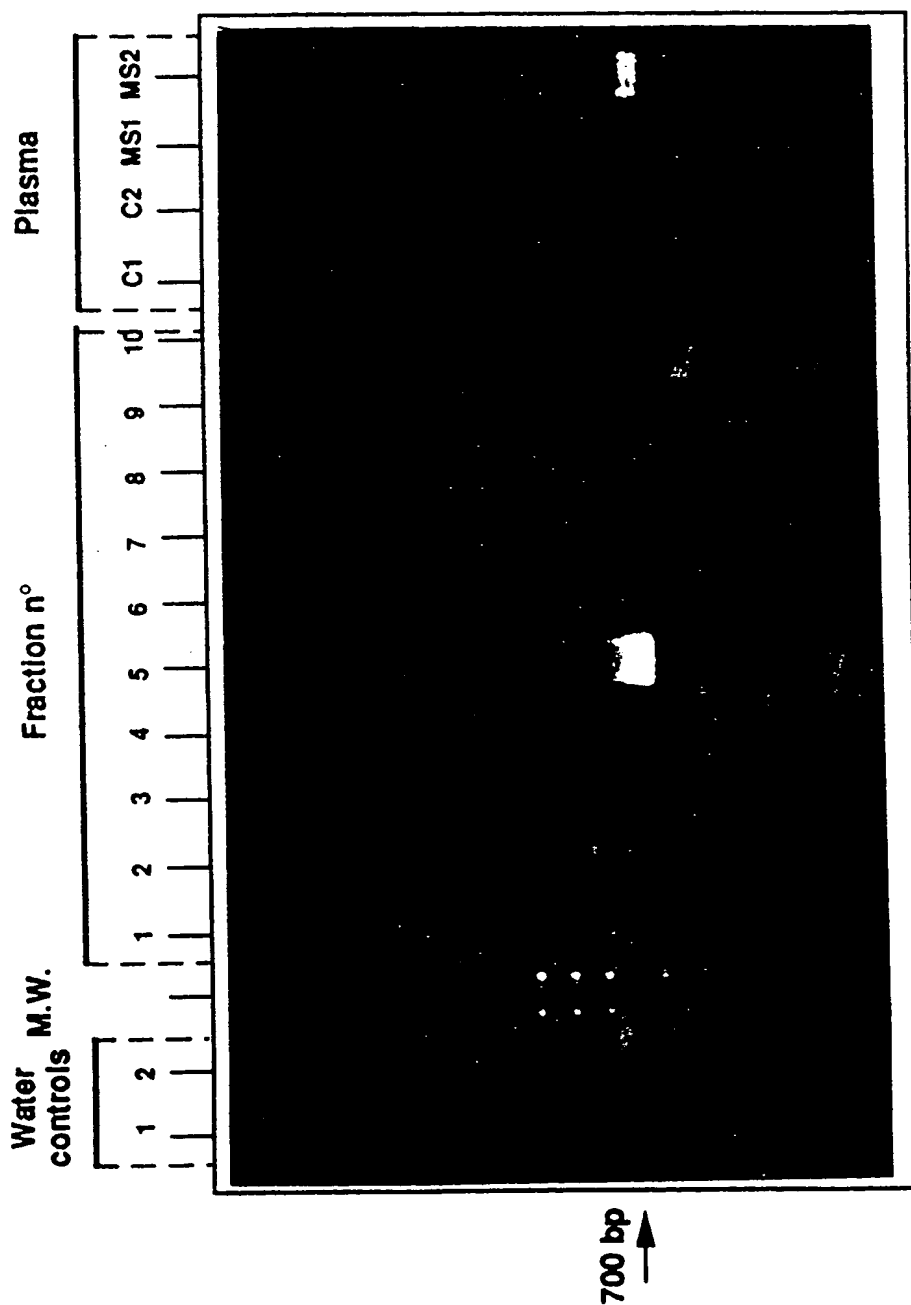
FIG 46

REVERSE-TRANSCRIPTASE &amp; RNase H

BNSDOCID: &lt;WO 9823755A1&gt;

53/69

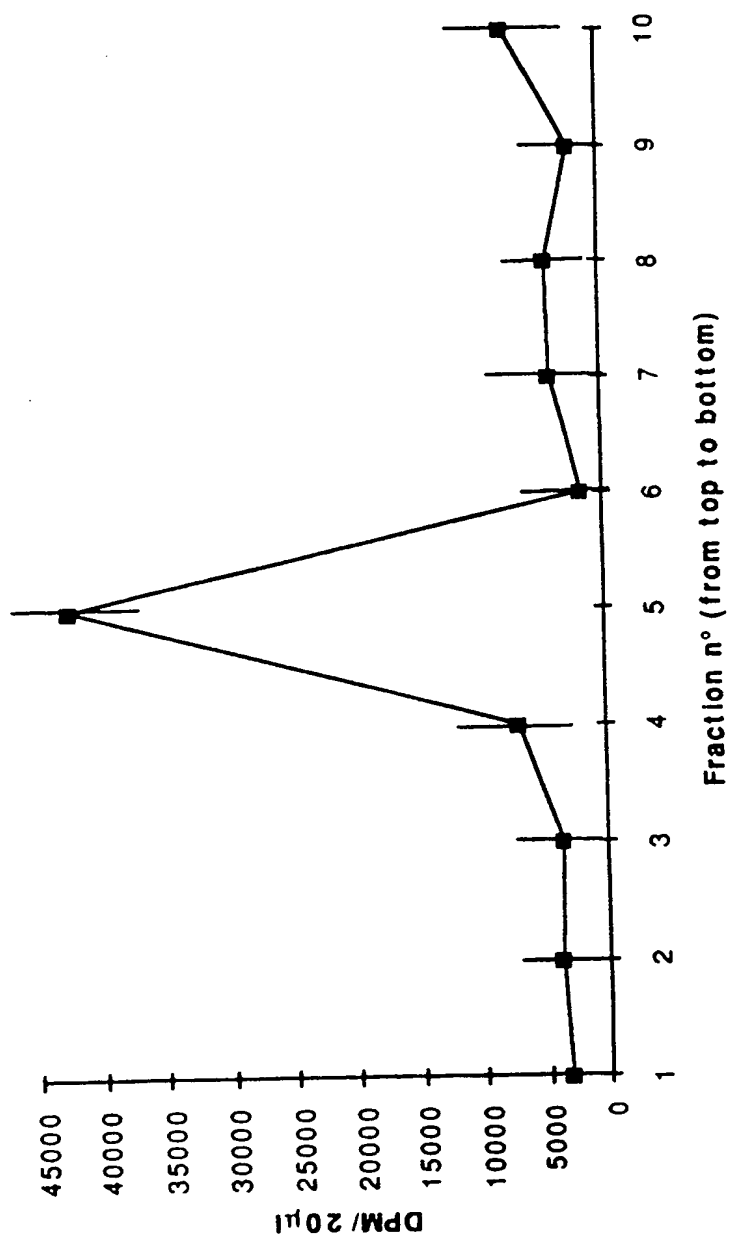
FIG 47A





54/69

FIG 47B



## FIG 48A 55/69

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
ATGATCCAGC	AGCAGGACNG	AGGGTGCCCG	GGGCAAGCGC	CAGCCCATGC	50
M I Q Q	Q D X	G C P	G Q A P	A H A	
CATCACCCCTC	ACAGAGCCCC	AGGTATGCTT	GACCATTGAG	GGTCAGAAGG	100
I T L	T E P Q	V C L	T I E	G Q K G	
GINACTGTCT	CCTGGACACT	GGCGNGCCT	TCTCAGTCTT	ACTTTCTGT	150
X C L	L D T	G G A F	S V L	L S C	
CCTGGACAAC	TGTCTCCAG	ATCTGTCACT	GTCCGAGGGG	TCCTAGGACA	200
P G Q L	S S R	S V T	V R G V	L G Q	
GCCAGTCACT	AGATACTTCT	CCCAGCCACT	AAGTTGTGAC	TGGGGAACCT	250
P V T	R Y F S	Q P L	S C D	W G T L	
TACTCTTCCC	ACATGCTTTT	CTAATTATGC	CTGAAAGCCC	CACCTCTCTG	300
L F P	H A F	L I M P	E S P	T L L	
TTGGGGAGAG	ACATTCTAGC	AAAAGCAGGG	GCCATTATAC	ATGTGAATAT	350
L G R D	I L A	K A G	A I I H	V N I	
AGGAGAAGGA	ACAACTGTTT	GTTGTCCCCT	GCTTGAGGAA	GGAATTAATC	400
G E G	T T V C	C P L	L E E	G I N P	
CTGAAGTCCG	GGCAACAGAA	GGACAATATG	GACAAGCAAA	GAATGCCCGT	450
E V R	A T E	G Q Y G	Q A K	N A R	
CCTGTTCAAG	TTAAACTAAA	GGATTCCACC	TCCTTTCCCT	ACCAAAGGCA	500
P V Q V	K L K	D S T	S F P Y	Q R Q	

## FIG 48B 56/69

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
GTACCCCCCTC	AGACCCGAGA	CCCAACAAGA	ACTCCAAAAG	ATTGTAAAGG	550
Y P L	R P E T	Q Q E	L Q K	I V K D	
ACCTAAAAGC	CCAAGGCCTA	GTAAAACCAA	GCAATAGCCC	TTGCAAGACT	600
L K A	Q G L	V K P S	N S P	C K T	
CCAATTTTAG	GAGTAAGGAA	ACCCAACGGA	CAGTGGAGGT	TAGTGCAAGA	650
P I L G	V R K	P N G	Q W R L	V Q E	
ACTCAGGATT	ATCAATGAGG	CTGTTGTTC	TCTATACCCA	GCTGTACCTA	700
L R I	I N E A	V V P	L Y P	A V P N	
ACCTTATAC	AGTGCTTTTC	CAAATACCAG	AGGAAGCAGA	GTGGTTTACA	750
P Y T	V L S	Q I P E	E A E	W F T	
GTCTGGACC	TTAAGGATGC	CTTTTCTGTC	ATCCCTGTAC	GTCTGACTC	800
V L D L	K D A	F F C	I P V R	P D S	
TCAATTCCTG	TTTGCCCTTG	AAGATCCTTT	GAACCCAACG	TCTCAACTCA	850
Q F L	F A F E	D P L	N P T	S Q L T	
CCTGGACTGT	TTTACCCCAA	GGGTTACAGG	ATAGCCCCCA	TCTATTTGGC	900
W T V	L P Q	G F R D	S P H	L F G	
CAGGCATTAG	CCCAAGACTT	GAGTCAATTC	TCATACCTGG	ACACTCTTGT	950
Q A L A	Q D L	S Q F	S Y L D	T L V	
CCTTCAGTAC	ATGGATGATT	TACTTTTAGT	CGCCCGTTCA	GAAACCTTGT	1000
L Q Y	M D D L	L L V	A R S	E T L C	

## FIG 48C 57/69

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
GCCATCAAGC	CACCCAAGAA	CTCTTAACIT	TCCTCACTAC	CTGTGGCTAC	1050
H Q A	T Q E	L L T F	L T T	C G Y	
AAGGTTTCCA	AACCAAAGGC	TCGGCTCTGC	TCACAGGAGA	TTAGATACTN	1100
K V S K	P K A	R L C	S Q E I	R Y X	
AGGGCTAAAA	TTATCCAAAG	GCACCAGGGC	CCTCAGTGAG	GAAOGTATOC	1150
G L K	L S K G	T R A	L S E	E R I Q	
AGCCTATACT	GGCTTATCCT	CATCCCAAAA	CCCTAAAGCA	ACTAAGAGGG	1200
P I L	A Y P	H P K T	L K Q	L R G	
TTCCTTGGCA	TAACAGGITT	CTGCCGAAAA	CAGATTCCCA	GGTACASCCC	1250
F L G I	T G F	C R K	Q I P R	Y X P	
AATAGCCAGA	CCATTATATA	CACTAATTAN	GGAAACTCAG	AAAGCCAATA	1300
I A R	P L Y T	L I X	E T Q	K A N T	
CCTATTTAGT	AAGATGGACA	CCTACAGAAG	TGGCTTTCCA	GGCCCTAAAG	1350
Y L V	R W T	P T E V	A F Q	A L K	
AAGGCCCTAA	CCCAAGCCCC	AGTGTTCAGC	TTGCCAACAG	GGCAAGATTT	1400
K A L T	Q A P	V F S	L P T G	Q D F	
TTCTTTTATAT	GCCACAGAAA	AAACAGGAAT	AGCTCTAGGA	GTCCTTACGC	1450
S L Y	A T E K	T G I	A L G	V L T Q	
AGGTCTCAGG	GATGAGCTTG	CAACCCGTGG	TATACCTGAG	TAAGGAAATT	1500
V S G	M S L	Q P V V	Y L S	K E I	

## FIG 48D 58/69

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
GATGTAGTGG	CAAAGGGTTG	GCCTCATNGT	TTATGGGTAA	TGCNGGCAGT	1550
D V V A	K G W	P H X	L W V M	X A V	
AGCAGTCTNA	GTATCTGAAG	CAGTTAAAAT	AATACAGGGA	AGAGATCTTN	1600
A V X	V S E A	V K I	I Q G	R D L X	
CTGTGTGCAC	ATCTCATGAT	GIGAACGGCA	TACTSRCTGC	TAAAGGAGAC	1650
V W T	S H D	V N G I	L X A	K G D	
TTGTGGTTGT	CAGACAACCA	TTTACTTAAN	TAYCAGGCYY	TATTACTTGA	1700
L W L S	D N H	L L X	Y Q A L	L L E	
AGAGCCAGTG	CTGNGACTGC	GCACTTGTC	AACTCTTAAA	CCCAAACCTA	1750
E P V	L X L R	T C P	T L K	P K L M	
TGCTGCCAG	AAGGATCTTT	NIAGAGGTC	CCTTAGCCAA	CCCTGACCTC	1800
L P R	R I F	X E V P	L A N	P D L	
AACTATATAT	ATACTGATGG	AAGTTCGTTT	GTAGAAAAGG	GATTACAAAG	1850
N Y I Y	T D G	S S F	V E K G	L Q R	
GGNAGGATAT	NCCATAGGTG	TTAGTGATAA	AGCAGTACTT	GAAAGTAAGC	1900
X G Y	X I G V	S D K	A V L	E S K P	
CTCTTCCCCC	CCAGGGACCA	GCGCCCCCGT	TAGCAGAACT	AGTGGCACTG	1950
L P P	Q G P	A P P L	A E L	V A L	
ACCCCGGAG	CCTTAGAACT	TTGGAAAGGG	AGGAGGATAA	ATGIGTATAC	2000
T P R A	L E L	W K G	R R I N	V Y T	

59/69

## FIG 48E

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
<hr/>					
AGATAGCAAG	TATGCTTATC	TAATCCGAAA	TGCCCATGTT	GCAATATGGA	2050
D S K	Y A Y L	I R N	A H V	A I W K	
AAGAAAGGGA	GTTCTTAACC	TCTGGGGGAA	CCCCATTAA	ATACCACAAG	2100
E R E	F L T	S G G T	P I K	Y H K	
TTAATCATGG	AGTTATTGCA	CACAGTGCAA	AAACTCAAGG	AGGTGGAAGT	2150
L I M E	L L H	T V Q	K L K E	V E V	
CTTACACTGC	CAAAGCCATC	AGAAAAGGGA	AAGAGGGGAA	GAGCAGCATA	2200
L H C	Q S H Q	K R E	R G E	E Q H K	
AGTGGCTACA	GAGGCAAGGA	AAGACTAGCA	GAAAGGAAAG	AGAGAAAGAG	2250
W L Q	R Q G	K T S R	K E R	E K E	
ACAGAAAGTC	AGAGAGAGAG	AGAGGAAGAG	ACAGAGCACA	AAGAGGGAGT	2300
T E S Q	R E R	E E E	T E H K	E G V	
CAGACAGAGA	GAGAGACAGA	GAGTCAGAGA	GAAGGAAAGA	GAGAGAGGAA	2350
R E R	E R Q R	V R E	K E R	E R G R	
GAGACAAAGA	ATGA				2364
D K E	.				

60/69  
FIG 49A

Complement of 8/46-7 propre	GACTTGAGCC AGTCCTCATA CCTGGACATT CTTGTTCCTC AGTATGGGA	50
1 /46-7 propre	GACTTGAGCC AGTCCTCATA CCTGGACATT CTTGTTCCTC AGTATGGGA	50
Complement of c15 propre 46-7	GACTTGAGCC AGTCCTCATA CCTGGACATT CTTGTTCCTC AGTATGGGA	50
Consensus	GACTTGAGCC AGTCCTCATA CCTGGACATT CTTGTTCCTC AGTATGGGA	50
Complement of 8/46-7 propre	TGAAITTAATT ATAGCCACCC ATTCAGAAAC CTTGTGGCAT CAAGCCACCC	100
1 /46-7 propre	TGAAITTAATT ATAGCCACCC ATTCAGAAAC CTTGTGGCAT CAAGCCACCC	100
Complement of c15 propre 46-7	TGAAITTAATT ATAGCCACCC ATTCAGAAAC CTTGTGGCAT CAAGCCACCC	100
Consensus	TGAAITTAATT ATAGCCACCC ATTCAGAAAC CTTGTGGCAT CAAGCCACCC	100
Complement of 8/46-7 propre	AAGCGCTCTT AAATTTCCCTT GCTACCTGTG GCTCCAAACA AAGGGCTCAG	150
1 /46-7 propre	AAGCGCTCTT AAATTTCCCTT GCTACCTGTG GCTCCAAACA AAGGGCTCAG	150
Complement of c15 propre 46-7	AAGCGCTCTT AAATTTCCCTT GCTACCTGTG GCTCCAAACA AAGGGCTCAG	150
Consensus	AAGCGCTCTT AAATTTCCCTT GCTACCTGTG GCTCCAAACA AAGGGCTCAG	150
Complement of 8/46-7 propre	CTCTGCTCAC ACCAGGTTAA ATACTTAGGG CTAAAATTAT CCAAAGTCAC	200
1 /46-7 propre	CTCTGCTCAC ACCAGGTTAA ATACTTAGGG CTAAAATTAT CCAAAGTCAC	200
Complement of c15 propre 46-7	CTCTGCTCAC ACCAGGTTAA ATACTTAGGG CTAAAATTAT CCAAAGTCAC	200
Consensus	CTCTGCTCAC ACCAGGTTAA ATACTTAGGG CTAAAATTAT CCAAAGTCAC	200
Complement of 8/46-7 propre	CAGGGCCCTC AGAGAGGAAC GTATCCAGCG TATACTGGGT TATCCCATC	250
1 /46-7 propre	CAGGGCCCTC AGAGAGGAAC GTATCCAGCG TATACTGGGT TATCCCATC	250
Complement of c15 propre 46-7	CAGGGCCCTC AGAGAGGAAC GTATCCAGCG TATACTGGGT TATCCCATC	250
Consensus	CAGGGCCCTC AGAGAGGAAC GTATCCAGCG TATACTGGGT TATCCCATC	250
Complement of 8/46-7 propre	CCAAACCCCT AAAGCAACTA AGAGGGTTC TTGGCATATC AGCCTTCTGC	300
1 /46-7 propre	CCAAACCCCT AAAGCAACTA AGAGGGTTC TTGGCATATC AGCCTTCTGC	300
Complement of c15 propre 46-7	CCAAACCCCT AAAGCAACTA AGAGGGTTC TTGGCATATC AGCCTTCTGC	300
Consensus	CCAAACCCCT AAAGCAACTA AGAGGGTTC TTGGCATATC AGCCTTCTGC	300
Complement of 8/46-7 propre	CGAATATGGA TTCCCGATA CAGTGAAATA GCCAGGCCAT TATGTACATT	350
1 /46-7 propre	CGAATATGGA TTCCCGATA CAGTGAAATA GCCAGGCCAT TATGTACATT	350
Complement of c15 propre 46-7	CGAATATGGA TTCCCGATA CAGTGAAATA GCCAGGCCAT TATGTACATT	350
Consensus	CGAATATGGA TTCCCGATA CAGTGAAATA GCCAGGCCAT TATGTACATT	350
Complement of 8/46-7 propre	AGTTAAGGAA ACTCAGAAAG CCAATACCCA TATAGTAAGA TGGACACCTG	400
1 /46-7 propre	AGTTAAGGAA ACTCAGAAAG CCAATACCCA TATAGTAAGA TGGACACCTG	400
Complement of c15 propre 46-7	AGTTAAGGAA ACTCAGAAAG CCAATACCCA TATAGTAAGA TGGACACCTG	400
Consensus	AGTTAAGGAA ACTCAGAAAG CCAATACCCA TATAGTAAGA TGGACACCTG	400
Complement of 8/46-7 propre	AAACAGAAAT GGCTTTCCAG GCCCTAAAG	429
1 /46-7 propre	AAACAGAAAT GGCTTTCCAG GCCCTAAAG	429
Complement of c15 propre 46-7	AAACAGAAAT GGCTTTCCAG GCCCTAAAG	429
Consensus	AAACAGAAAT GGCTTTCCAG GCCCTAAAG	429

## FIG 49B

Trans of 1 /46-7 pr	DLSQSSYLDI LVLQYEDDLI IATHSETLWH QATQALLNFL ATCGSKQKAAH	50
Trans of Complement-2(8)	DLSQSSYLDI LVLQYEDDLI IATHSETLWH QATQALLNFL ATCGSKQKAAQ	50
Trans of Complement(5)	DLSQSSYLDI LVLQYEDDLI IATHSETLWH QATQALLNFL ATCGSKQKAAQ	50
Consensus	DLSQSSYLDI LVLQYEDDLI IATHSETLWH QATQALLNFL ATCGSKQKAAQ	50
Trans of 1 /46-7 pr	LCSQVKYLG LKLSKVIRAL REERIQRILA YPHPKTKQL RFLGLIHAFC	100
Trans of Complement-2	LCSQVKYLG LKLSKVIRAL REERIQRILD YPHPKTKQL RFLGLIHAFC	100
Trans of Complement	LCSQVKYLG LKLSKVIRAL REERIQRILA YPHPKTKQL RFLGLIHAFC	100
Consensus	LCSQVKYLG LKLSKVIRAL REERIQRILA YPHPKTKQL RFLGLIHAFC	100
Trans of 1 /46-7 pr	RIWIFRYSEI ARPLCTLKE TQKANTHIVR WTPETEVAFQ ALK	143
Trans of Complement-2	RIWIFRYSEI ARPLCTLKE TQKANTHIVR WTPETEVAFQ ALK	143
Trans of Complement	RIWIFRYSEI ARPLCTLKE TQKANTHIVR WTPETEVAFQ ALK	143
Consensus	RIWIFRYSEI ARPLCTLKE TQKANTHIVR WTPETEVAFQ ALK	143

## FIG 50B

Trans of c143 propr	DLSQSSYLDX LVLRYMDDL LATHSETLCH QATQALLNFL ATCGYKVSKE	50
Trans of 42/68-1 pr	DLSQSSYLDI LVLRYMDDL LATHSETLCH QATQALLNFL ATCGYKVSKE	50
Trans of 41/68-1 pr	DLSQSSYLDI LVLRYMDDL LATHSETLCH QATQALLNFL ATCGYKVSKE	50
Consensus	DLSQSSYLDI LVLRYMDDL LATHSETLCH QATQALLNFL ATCGYKVSKE	50
Trans of c143 propr	KAQLCSQQVK YLGLKLSKGT RTLSEERIOP ILGYPHPKTL KQLTAFLGIT	100
Trans of 42/68-1 pr	KAQLCSQQVK YLGLKLSKGT RTLSEERIOP ILGYPHPKTL KQLTAFLGIT	100
Trans of 41/68-1 pr	KAQLCSQQVK YLGLKLSKGT RTLSEERIOP ILGYPHPKTL KQLTAFLGIT	100
Consensus	KAQLCSQQVK YLGLKLSKGT RTLSEERIOP ILGYPHPKTL KQLTAFLGIT	100
Trans of c143 propr	GFCQIWIIPRY SKIARPLNTR IKETQKANTH LVRWTFEAEV AFQALK	146
Trans of 42/68-1 pr	GFCQIWIIPRY SKIARPLNTR IKETQKANTH LVRWTFEAEV AFQALK	146
Trans of 41/68-1 pr	GFCQIWIIPRY SKIARPLNTR IKETQKANTH LVRWTFEAEV AFQALK	146
Consensus	GFCQIWIIPRY SKIARPLNTR IKETQKANTH LVRWTFEAEV AFQALK	146



62/69

## FIG 50A

41/68-1 propre	GACTTGAGCC AGTCATCATA CCTGGACAMT CTTGTCCCTTC GGTACATGGA	50
c143 propre 68-1	GACTTGAGCC AGTCATCATA CCTGGACAMT CTTGTCCCTTC GGTACATGGA	50
42/68-1 propre	GACTTGAGCC AGTCATCATA CCTGGACAMT CTTGTCCCTTC GGTACATGGA	50
Consensus	GACTTGAGCC AGTCATCATA CCTGGACAMT CTTGTCCCTTC GGTACATGGA	50
41/68-1 propre	TGATTTACTTT TTAGCCACCC ATTCAGAAAC CTTGTGCCAT CAAGCCACCC	100
c143 propre 68-1	TGATTTACTTT TTAGCCACCC ATTCAGAAAC CTTGTGCCAT CAAGCCACCC	100
42/68-1 propre	TGATTTACTTT TTAGCCACCC ATTCAGAAAC CTTGTGCCAT CAAGCCACCC	100
Consensus	TGATTTACTTT TTAGCCACCC ATTCAGAAAC CTTGTGCCAT CAAGCCACCC	100
41/68-1 propre	AAGCACTCTTT AAATTTCCTT GCTACCTGTG GCTACAAGGT TTCCAAACCA	150
c143 propre 68-1	AAGCACTCTTT AAATTTCCTT GCTACCTGTG GCTACAAGGT TTCCAAACCA	150
42/68-1 propre	AAGCACTCTTT AAATTTCCTT GCTACCTGTG GCTACAAGGT TTCCAAACCA	150
Consensus	AAGCACTCTTT AAATTTCCTT GCTACCTGTG GCTACAAGGT TTCCAAACCA	150
41/68-1 propre	AAGGCTCAGC TCTGCTCACA GCAGGTTAAA TACTTAGGGC TAAAATTATC	200
c143 propre 68-1	AAGGCTCAGC TCTGCTCACA GCAGGTTAAA TACTTAGGGC TAAAATTATC	200
42/68-1 propre	AAGGCTCAGC TCTGCTCACA GCAGGTTAAA TACTTAGGGC TAAAATTATC	200
Consensus	AAGGCTCAGC TCTGCTCACA GCAGGTTAAA TACTTAGGGC TAAAATTATC	200
41/68-1 propre	CAAAGGCACC AGAACCCCTCA GTGAGGAACG TATCCAGCCT ATACTGGGTT	250
c143 propre 68-1	CAAAGGCACC AGAACCCCTCA GTGAGGAACG TATCCAGCCT ATACTGGGTT	250
42/68-1 propre	CAAAGGCACC AGAACCCCTCA GTGAGGAACG TATCCAGCCT ATACTGGGTT	250
Consensus	CAAAGGCACC AGAACCCCTCA GTGAGGAACG TATCCAGCCT ATACTGGGTT	250
41/68-1 propre	ATCCTCATCC CAAAACCCCTA AAGCAACTAA CAGCGTTCCT TGGCATAACA	300
c143 propre 68-1	ATCCTCATCC CAAAACCCCTA AAGCAACTAA CAGCGTTCCT TGGCATAACA	300
42/68-1 propre	ATCCTCATCC CAAAACCCCTA AAGCAACTAA CAGCGTTCCT TGGCATAACA	300
Consensus	ATCCTCATCC CAAAACCCCTA AAGCAACTAA CAGCGTTCCT TGGCATAACA	300
41/68-1 propre	GGTTTCTGCC AAATATGGAT TCCCAGGTAC AGCAAAATAG CCAGACCATT	350
c143 propre 68-1	GGTTTCTGCC AAATATGGAT TCCCAGGTAC AGCAAAATAG CCAGACCATT	350
42/68-1 propre	GGTTTCTGCC AAATATGGAT TCCCAGGTAC AGCAAAATAG CCAGACCATT	350
Consensus	GGTTTCTGCC AAATATGGAT TCCCAGGTAC AGCAAAATAG CCAGACCATT	350
41/68-1 propre	AAATACACGA ATTAAGGAAA CTCAAAAAGC CATTACCCAT TTAGTAAGAT	400
c143 propre 68-1	AAATACACGA ATTAAGGAAA CTCAAAAAGC CATTACCCAT TTAGTAAGAT	400
42/68-1 propre	AAATACACGA ATTAAGGAAA CTCAAAAAGC CATTACCCAT TTAGTAAGAT	400
Consensus	AAATACACGA ATTAAGGAAA CTCAAAAAGC CATTACCCAT TTAGTAAGAT	400
41/68-1 propre	GGACAATCTGA AGCAGAAGTG GCTTTCCAGG CCCTAAAG	438
c143 propre 68-1	GGACAATCTGA AGCAGAAGTG GCTTTCCAGG CCCTAAAG	438
42/68-1 propre	GGACAATCTGA AGCAGAAGTG GCTTTCCAGG CCCTAAAG	438
Consensus	GGACAATCTGA AGCAGAAGTG GCTTTCCAGG CCCTAAAG	438

## FIG 51A

MSRV pol	ATTATGCCTG	AAAGCCCCAC	TCCCTTGTTA	GGGAGAGACA	TTTTAGCAAA	50
cons ADN 1,5,8	-----	-----	-----	-----	-----	
Consensus	ATTATGCCTG	AAAGCCCCAC	TCCCTTGTTA	GGGAGAGACA	TTTTAGCAAA	50
MSRV pol	AGCAGGGGCC	ATTATACACC	TGAACATAGG	AAAAGGAATA	CCCATTTGCT	100
cons ADN 1,5,8	-----	-----	-----	-----	-----	
Consensus	AGCAGGGGCC	ATTATACACC	TGAACATAGG	AAAAGGAATA	CCCATTTGCT	100
MSRV pol	GTCCCCTGCT	TGAGGAAGGA	ATTAATCCTG	AAGTCTGGGC	AATAGAAGGA	150
cons ADN 1,5,8	-----	-----	-----	-----	-----	
Consensus	GTCCCCTGCT	TGAGGAAGGA	ATTAATCCTG	AAGTCTGGGC	AATAGAAGGA	150
MSRV pol	CAATATGGAC	AAGCAAAGAA	TGCCCGTCCT	GTTCAGTTA	AACTAAAGGA	200
cons ADN 1,5,8	-----	-----	-----	-----	-----	
Consensus	CAATATGGAC	AAGCAAAGAA	TGCCCGTCCT	GTTCAGTTA	AACTAAAGGA	200
MSRV pol	TTCTGCCTCC	TTTCCTTACC	AAAGGAAGTA	CCCTCTTAGA	CCCGAGGCC	250
cons ADN 1,5,8	-----	-----	-----	-----	-----	
Consensus	TTCTGCCTCC	TTTCCTTACC	AAAGGAAGTA	CCCTCTTAGA	CCCGAGGCC	250
MSRV pol	TACAAGGANC	TCAAAGATT	GTTAAGGACC	TAAAAGCCCA	AGGCCTAGTA	300
cons ADN 1,5,8	-----	-----	-----	-----	-----	
Consensus	TACAAGGANC	TCAAAGATT	GTTAAGGACC	TAAAAGCCCA	AGGCCTAGTA	300
MSRV pol	AAACCATGCA	GTAGCCCCTG	CAATACTCCA	ATTTTAGGAG	TAAGGAAACC	350
cons ADN 1,5,8	-----	-----	-----	-----	-----	
Consensus	AAACCATGCA	GTAGCCCCTG	CAATACTCCA	ATTTTAGGAG	TAAGGAAACC	350
MSRV pol	CAACGGACAG	TGGAGGTTAG	TGCAAGATCT	CAGGATTATT	AATGAGGCTG	400
cons ADN 1,5,8	-----	-----	-----	-----	-----	
Consensus	CAACGGACAG	TGGAGGTTAG	TGCAAGATCT	CAGGATTATT	AATGAGGCTG	400
MSRV pol	TTTTTCCTCT	ATACCCAGCT	GTATCTAGCC	CTTATACTCT	GCTTTCCCTA	450
cons ADN 1,5,8	-----	-----	-----	-----	-----	
Consensus	TTTTTCCTCT	ATACCCAGCT	GTATCTAGCC	CTTATACTCT	GCTTTCCCTA	450
MSRV pol	ATACCAGAGG	AAGCAGAGTG	GTTTACAGTC	CTGGACCTTA	AGGATGCCTT	500
cons ADN 1,5,8	-----	-----	-----	-----	-----	
Consensus	ATACCAGAGG	AAGCAGAGTG	GTTTACAGTC	CTGGACCTTA	AGGATGCCTT	500
MSRV pol	TTTCTGCATC	CCTGTACGTC	CTGACTCTCA	ATTCTTGTTT	GCCTTTGAAG	550
cons ADN 1,5,8	-----	-----	-----	-----	-----	
Consensus	TTTCTGCATC	CCTGTACGTC	CTGACTCTCA	ATTCTTGTTT	GCCTTTGAAG	550
MSRV pol	ATCCTTTGAA	CCCAACGTCT	CAACTCACCT	GGACTGTTTT	ACCCCAAGGG	600
cons ADN 1,5,8	-----	-----	-----	-----	-----	
Consensus	ATCCTTTGAA	CCCAACGTCT	CAACTCACCT	GGACTGTTTT	ACCCCAAGGG	600
MSRV pol	TTCAGGGATA	GCCCCATCT	ATTTGGCCAG	GCATTAGCCC	ANGACTTGAG	650
cons ADN 1,5,8	-----	-----	-----	-----	-----	
Consensus	TTCAGGGATA	GCCCCATCT	ATTTGGCCAG	GCATTAGCCC	ANGACTTGAG	650
MSRV pol	TCATTTCTCA	TACCTGGACA	TCTTGTGCT	TCAGTACGTC	GATGAMTTAC	700
cons ADN 1,5,8	-----	-----	-----	-----	-----	
Consensus	TCATTTCTCA	TACCTGGACA	TCTTGTGCT	TCAGTACGTC	GATGAMTTAC	700
MSRV pol	TTTATAGTCC	CCCTTCAGAA	ACCTTGTGTC	ATCAAGCCAC	CCAAGATCTC	750
cons ADN 1,5,8	-----	-----	-----	-----	-----	
Consensus	TTTATAGTCC	CCCTTCAGAA	ACCTTGTGTC	ATCAAGCCAC	CCAAGATCTC	750
MSRV pol	TTTAAATTTC	TACTTACCTG	TGGCTACAAG	GTTCCTCAAC	TAAGGGCTGC	800
cons ADN 1,5,8	-----	-----	-----	-----	-----	
Consensus	TTTAAATTTC	TACTTACCTG	TGGCTACAAG	GTTCCTCAAC	TAAGGGCTGC	800

## FIG 51A (cont.)

MSRV pol	GCTCTGCTCA CAGCAGTTTA TATACTTAGG GCTAAAATTA TCCAAAGGCA	850
cons ADN 1,5,8	SCTCTGCTCA CAGCAGTTTA TATACTTAGG GCTAAAATTA TCCAAAGGCR	199
Consensus	SCTCTGCTCA CAGCAGTTTA TATACTTAGG GCTAAAATTA TCCAAAGGCR	850
MSRV pol	CCAGGGCCCT CAGTGAGGAA CGTATCCAGC TATACTGGC TTATCCCAT	900
cons ADN 1,5,8	CCAGGGCCCT CAGTGAGGAA CGTATCCAGC TATACTGGM TTATCCCAT	249
Consensus	CCAGGGCCCT CAGTGAGGAA CGTATCCAGC TATACTGGM TTATCCCAT	900
MSRV pol	CCCAAAACCC TAAAGCAACT AAGAGGGTTC CTTGGCATAA CAGGTTTCTG	950
cons ADN 1,5,8	CCCAAAACCC TAAAGCAACT AAGAGGGTTC CTTGGCATAA CAGGTTTCTG	299
Consensus	CCCAAAACCC TAAAGCAACT AAGAGGGTTC CTTGGCATAA CAGGTTTCTG	950
MSRV pol	CCGAATATCG ATTCCCGGT ACACCCCAAT AGCCAGGCCA TTATGTACAC	1000
cons ADN 1,5,8	CCGAATATCG ATTCCCGGT ACAGSYGHAAT AGCCAGGCCA TTATGTACAT	349
Consensus	CCGAATATCG ATTCCCGGT ACAGSYMAAT AGCCAGGCCA TTATGTACAT	1000
MSRV pol	TAAATTAAGGA AACTCAGAAA GCCAATACCT ATTTAGTAAG ATGGACACCT	1050
cons ADN 1,5,8	TADYTAAGGA AACTCAGAAA GCCAATACCC ATTTAGTAAG ATGGACACCT	399
Consensus	TADYTAAGGA AACTCAGAAA GCCAATACCT ATTTAGTAAG ATGGACACCT	1050
MSRV pol	---ACAGAAG TGGCTTTCCA GGCCCTAAAG AAGGCCCTAA CCCAAGCCCC	1097
cons ADN 1,5,8	GAFACAGAAG TGGCTTTCCA GGCCCTAAAG AAGGCCCTAA CCCAAGCCCC	429
Consensus	GAFACAGAAG TGGCTTTCCA GGCCCTAAAG AAGGCCCTAA CCCAAGCCCC	1100
MSRV pol	AGTGTTCAGC TTGCCAACAG GGCAAGATTT TTCTTTATAT GCCACAGAAA	1147
cons ADN 1,5,8	-----	429
Consensus	AGTGTTCAGC TTGCCAACAG GGCAAGATTT TTCTTTATAT GCCACAGAAA	1150
MSRV pol	AAACAGGAAT AGCTCTAGGA GTCCTTACGC AGGTCTCAGG GATGAGCTTG	1197
cons ADN 1,5,8	-----	429
Consensus	AAACAGGAAT AGCTCTAGGA GTCCTTACGC AGGTCTCAGG GATGAGCTTG	1200
MSRV pol	CAACCCGTGG TATACCTGAG TAAGGAAATT GATGTAGTGG CAAAGGGTTG	1247
cons ADN 1,5,8	-----	429
Consensus	CAACCCGTGG TATACCTGAG TAAGGAAATT GATGTAGTGG CAAAGGGTTG	1250
MSRV pol	GCCTCATTTG TTATGGGTAA TGGCGGCAGT AGCAGTCTTA GTATCTGAAG	1297
cons ADN 1,5,8	-----	429
Consensus	GCCTCATTTG TTATGGGTAA TGGCGGCAGT AGCAGTCTTA GTATCTGAAG	1300
MSRV pol	CAGTTAAAT AATACAGGGA AGAGATCTTA CTGTGTGGAC ATCTCATGAT	1347
cons ADN 1,5,8	-----	429
Consensus	CAGTTAAAT AATACAGGGA AGAGATCTTA CTGTGTGGAC ATCTCATGAT	1350
MSRV pol	GTGAACGGCA TACTCACTGC TAAAGGAGAC TTGTGGTTGT CAGACAACCA	1397
cons ADN 1,5,8	-----	429
Consensus	GTGAACGGCA TACTCACTGC TAAAGGAGAC TTGTGGTTGT CAGACAACCA	1400
MSRV pol	TTTACTTAAT TATCAGGCTC TATTACTTGA AGAGCCAGTG CTGAGACTGC	1447
cons ADN 1,5,8	-----	429
Consensus	TTTACTTAAT TATCAGGCTC TATTACTTGA AGAGCCAGTG CTGAGACTGC	1450
MSRV pol	GCACTTGTGC AACTCTTAAA CCCGCCACAT TTCTTCCAGA CAATGAAGAA	1497
cons ADN 1,5,8	-----	429
Consensus	GCACTTGTGC AACTCTTAAA CCCGCCACAT TTCTTCCAGA CAATGAAGAA	1500
MSRV pol	AAGATAGAAC ATAACGTGCA ACAAGTAATT GCTCAAACCT ATGCTGCTCG	1547
cons ADN 1,5,8	-----	429
Consensus	AAGATAGAAC ATAACGTGCA ACAAGTAATT GCTCAAACCT ATGCTGCTCG	1550
MSRV pol	AGGGGACCTT CTAGAGGTTT CCTTGACTGA TCCCGACCTC AACTTGTATA	1597
cons ADN 1,5,8	-----	429
Consensus	AGGGGACCTT CTAGAGGTTT CCTTGACTGA TCCCGACCTC AACTTGTATA	1600

65/69

## FIG 51 B

Trans of MSRV pol cons prot 1,5,8	IMPESPTPLL GRDILAKAGA IHLNIGKGI PICCPLEEG INPEVWAIEG	50
Consensus	.....	50
Trans of MSRV pol cons prot 1,5,8	QYGOAKNARP VQVKLKDSAS FPYQRKYPLR PEALQGXQKI VKDLKAQGLV	100
Consensus	.....	100
Trans of MSRV pol cons prot 1,5,8	KPCSSPCNTP ILGVRKPNGQ WRLVQDLRII NEAVFPLYPA VSSPYTLLSL	150
Consensus	.....	150
Trans of MSRV pol cons prot 1,5,8	IPEEAWEFTV LDLKDAFFCI FVRPDSQFLF AFEDPLNPTS QLTWIVLPQG	200
Consensus	.....	200
Trans of MSRV pol cons prot 1,5,8	FRDSPHLFGQ ALACDLSQSS YLDLVLQYV DDLILVARSE TLHQATQEH	250
Consensus	.....	36
	.....	250
Trans of MSRV pol cons prot 1,5,8	TLFLPTCGSK VSKKALCS QELRYLGLKL SKVTRALSEE RIQHTLAYPH	300
Consensus	.....	83
	.....	300
Trans of MSRV pol cons prot 1,5,8	PKTLKQLRGF LGITFCRQK IPRYTHIARP LMTLIRETQK ANIVLVRWTF	350
Consensus	.....	133
	.....	350
Trans of MSRV pol cons prot 1,5,8	TEVAFQALK KALTQAPVFS LPTGQDFSLY ATEKTGIALG VLTQVSGMSL	399
Consensus	.....	143
	.....	400
Trans of MSRV pol cons prot 1,5,8	QPVVYLSKEI DVVAKGWEHC LWMMAVAVL VSEAVKIIQG RDLTWVTSHD	449
Consensus	.....	143
	.....	450
Trans of MSRV pol cons prot 1,5,8	VNGILTAKGD LWLSDNHLIN YQALLLEFPV LRLRTCATLK PATFLPINEE	499
Consensus	.....	143
	.....	500
Trans of MSRV pol cons prot 1,5,8	KIEHNCQQVI AQTAAARGDL LEVPLTDPDL NLYTDGSSLA EKGLRKAGYA	549
Consensus	.....	143
	.....	550
Trans of MSRV pol cons prot 1,5,8	VISINGILES NRLTPGTSAH LAELIALTWA LELGEGKRVN IYSDSKYAYL	599
Consensus	.....	143
	.....	600
Trans of MSRV pol cons prot 1,5,8	VLHAHAATWR EREFLTSEGT PINHQEAIIR LLLAVQKPKE VAVLHCQGHQ	649
Consensus	.....	143
	.....	650
Trans of MSRV pol cons prot 1,5,8	EEEEEREIEGN RQADIEAKKA ARQDSPLEML IEGP	683
Consensus	.....	143
	.....	684

66/69

## FIG 52 A

MSRV pol	ATTATGCCTG AAAGCCCCAC TCCCTTGTTA GGGAGAGACA TTTTAGCAAA	50
cons ADN 41.42.43	-----	
Consensus	ATTATGCCTG AAAGCCCCAC TCCCTTGTTA GGGAGAGACA TTTTAGCAAA	50
MSRV pol	AGCAGGGGCC ATTATACACC TGAACATAGG AAAAGGAATA CCCATTGCT	100
cons ADN 41.42.43	-----	
Consensus	AGCAGGGGCC ATTATACACC TGAACATAGG AAAAGGAATA CCCATTGCT	100
MSRV pol	GTCCCTGCT TGAGGAAGGA ATTAATCCTG AAGTCTGGGC AATAGAAGGA	150
cons ADN 41.42.43	-----	
Consensus	GTCCCTGCT TGAGGAAGGA ATTAATCCTG AAGTCTGGGC AATAGAAGGA	150
MSRV pol	CAATATGGAC AAGCAAAGAA TGCCCGTCT GTTCAAGTTA AACTAAAGGA	200
cons ADN 41.42.43	-----	
Consensus	CAATATGGAC AAGCAAAGAA TGCCCGTCT GTTCAAGTTA AACTAAAGGA	200
MSRV pol	TTCTGCCTCC TTTCCTTACC AAAGGAAGTA CCTCTTAGA CCCGAGGCC	250
cons ADN 41.42.43	-----	
Consensus	TTCTGCCTCC TTTCCTTACC AAAGGAAGTA CCTCTTAGA CCCGAGGCC	250
MSRV pol	TACAAGGANC TCAAAGATT GTTAAGGACC TAAAAGCCCA AGGCTAGTA	300
cons ADN 41.42.43	-----	
Consensus	TACAAGGANC TCAAAGATT GTTAAGGACC TAAAAGCCCA AGGCTAGTA	300
MSRV pol	AAACCATGCA GTAGCCCCTG CAATACTCCA ATTTTAGGAG TAAGGAAACC	350
cons ADN 41.42.43	-----	
Consensus	AAACCATGCA GTAGCCCCTG CAATACTCCA ATTTTAGGAG TAAGGAAACC	350
MSRV pol	CAACGGACAG TGGAGGTTAG TGCAAGATCT CAGGATTATT AATGAGGCTG	400
cons ADN 41.42.43	-----	
Consensus	CAACGGACAG TGGAGGTTAG TGCAAGATCT CAGGATTATT AATGAGGCTG	400
MSRV pol	TTTTTCCTCT ATACCCAGCT GTATCTAGCC CTATACTCT GCTTTCCTTA	450
cons ADN 41.42.43	-----	
Consensus	TTTTTCCTCT ATACCCAGCT GTATCTAGCC CTATACTCT GCTTTCCTTA	450
MSRV pol	ATACCAGAGG AAGCAGAGTG GTTACAGTC CTGGACCTTA AGGATGCCTT	500
cons ADN 41.42.43	-----	
Consensus	ATACCAGAGG AAGCAGAGTG GTTACAGTC CTGGACCTTA AGGATGCCTT	500
MSRV pol	TTTCTGCATC CCTGTACGTC CTGACTCTCA ATTCTTGTTT GCCTTTGAAG	550
cons ADN 41.42.43	-----	
Consensus	TTTCTGCATC CCTGTACGTC CTGACTCTCA ATTCTTGTTT GCCTTTGAAG	550
MSRV pol	ATCCTTTGAA CCCAAGCTCT CAACTCACCT GGACTGTTT ACCCCAAGGG	600
cons ADN 41.42.43	-----	
Consensus	ATCCTTTGAA CCCAAGCTCT CAACTCACCT GGACTGTTT ACCCCAAGGG	600
MSRV pol	TTCAGGGATA GCCCCATCT ATTGGCCAG GCATTAGCCC AAGACTTGAG	650
cons ADN 41.42.43	-----	
Consensus	TTCAGGGATA GCCCCATCT ATTGGCCAG GCATTAGCCC AAGACTTGAG	650
MSRV pol	TCAATTCACA TACCTGGACA TCTTGTCCT TCGTACCTG GATGATTAC	700
cons ADN 41.42.43	-----	
Consensus	TCAATTCACA TACCTGGACA TCTTGTCCT TCGTACCTG GATGATTAC	700
MSRV pol	TTTTAGGCTC CCTTCAGAA ACCTTGTCCT ATCAAGCCAC CCAAGACTC	750
cons ADN 41.42.43	-----	
Consensus	TTTTAGGCTC CCTTCAGAA ACCTTGTCCT ATCAAGCCAC CCAAGACTC	750
MSRV pol	TTAACTTTC TCACTACCTG TGGCTACAAG GTTCCAAAC CAAAGGCTC	800
cons ADN 41.42.43	-----	
Consensus	TTAACTTTC TCACTACCTG TGGCTACAAG GTTCCAAAC CAAAGGCTC	800

67/69

## FIG 52 A (cont.)

MSRV pol	GCTCTGCTCA CAGCAGTTTA TATACTTAGG GCTAAAATTA TCCAAAGGCA	850
cons ADN 41,42,43	GCTCTGCTCA CAGCAGTTTA TATACTTAGG GCTAAAATTA TCCAAAGGCA	208
Consensus	GCTCTGCTCA CAGCAGTTTA TATACTTAGG GCTAAAATTA TCCAAAGGCA	850
MSRV pol	CCAGGCCCCT CAGTGAGGAA CGTATCCAGC CTATACTGGC TTATCCTCAT	900
cons ADN 41,42,43	CCAGGCCCCT CAGTGAGGAA CGTATCCAGC CTATACTGGC TTATCCTCAT	258
Consensus	CCAGGCCCCT CAGTGAGGAA CGTATCCAGC CTATACTGGC TTATCCTCAT	900
MSRV pol	CCCAAACCC TAAAGCAACT AAGAGGTTTC CTGGGCATAA CAGGTTTCTG	950
cons ADN 41,42,43	CCCAAACCC TAAAGCAACT AAGAGGTTTC CTGGGCATAA CAGGTTTCTG	308
Consensus	CCCAAACCC TAAAGCAACT AAGAGGTTTC CTGGGCATAA CAGGTTTCTG	950
MSRV pol	CCCAAAACAG ATTCCAGGT ACACCCCAAT AGCCAGACCA TTATATACAC	1000
cons ADN 41,42,43	CCCAAAACAG ATTCCAGGT ACACCAARFT AGCCAGACCA TTATATACAC	358
Consensus	CCCAAAACAG ATTCCAGGT ACACCCMRRT AGCCAGACCA TTATATACAC	1000
MSRV pol	TAATTAGGGA AACTCAHAAA GCCARTACCT ATTTAGTAAG ATGGACAACCT	1050
cons ADN 41,42,43	GAATTAGGGA AACTCAHAAA GCCARTACCC ATTTAGTAAG ATGGACAACCT	408
Consensus	KAATTAGGGA AACTCAHAAA GCCARTACCT ATTTAGTAAG ATGGACAACCT	1050
MSRV pol	--A CAGAAG TGGCTTTCCA GGCCTTAAAG AAGGCCCTAA CCCAAGCCCC	1097
cons ADN 41,42,43	GAAC CAGAAG TGGCTTTCCA GGCCTTAAAG -----	438
Consensus	GAAC CAGAAG TGGCTTTCCA GGCCTTAAAG AAGGCCCTAA CCCAAGCCCC	1100
MSRV pol	AGTGTTCAGC TTGCCAACAG GGCAAGATTT TTCTTTATAT GCCACAGAAA	1147
cons ADN 41,42,43	-----	438
Consensus	AGTGTTCAGC TTGCCAACAG GGCAAGATTT TTCTTTATAT GCCACAGAAA	1150
MSRV pol	AAACAGGAAT AGCTCTAGGA GTCCTTACGC AGGTCTCAGG GATGAGCTTG	1197
cons ADN 41,42,43	-----	438
Consensus	AAACAGGAAT AGCTCTAGGA GTCCTTACGC AGGTCTCAGG GATGAGCTTG	1200
MSRV pol	CAACCCGTGG TATACCTGAG TAAGGAAATT GATGTAGTGG CAAAGGGTTG	1247
cons ADN 41,42,43	-----	438
Consensus	CAACCCGTGG TATACCTGAG TAAGGAAATT GATGTAGTGG CAAAGGGTTG	1250
MSRV pol	GCCTCATTGT TTATGGGTAA TGGCGGCAGT AGCAGTCTTA GTATCTGAAG	1297
cons ADN 41,42,43	-----	438
Consensus	GCCTCATTGT TTATGGGTAA TGGCGGCAGT AGCAGTCTTA GTATCTGAAG	1300
MSRV pol	CAGTTAAAT AATACAGGGA AGAGATCTTA CTGTGTGGAC ATCTCATGAT	1347
cons ADN 41,42,43	-----	438
Consensus	CAGTTAAAT AATACAGGGA AGAGATCTTA CTGTGTGGAC ATCTCATGAT	1350
MSRV pol	GTGAACGGCA TACTCACTGC TAAAGGAGAC TTGTGGTTGT CAGACAACCA	1397
cons ADN 41,42,43	-----	438
Consensus	GTGAACGGCA TACTCACTGC TAAAGGAGAC TTGTGGTTGT CAGACAACCA	1400
MSRV pol	TTTACTTAAT TATCAGGCTC TATTACTTGA AGAGCCAGTG CTGAGACTGC	1447
cons ADN 41,42,43	-----	438
Consensus	TTTACTTAAT TATCAGGCTC TATTACTTGA AGAGCCAGTG CTGAGACTGC	1450
MSRV pol	GCACTTGTGC AACTCTTAAA CCCGCCACAT TTCTTCCAGA CAATGAAGAA	1497
cons ADN 41,42,43	-----	438
Consensus	GCACTTGTGC AACTCTTAAA CCCGCCACAT TTCTTCCAGA CAATGAAGAA	1500
MSRV pol	AAGATAGAAC ATAACGTGCA ACAAGTAATT GCTCAAACCT ATGCTGCTCG	1547
cons ADN 41,42,43	-----	438
Consensus	AAGATAGAAC ATAACGTGCA ACAAGTAATT GCTCAAACCT ATGCTGCTCG	1550
MSRV pol	AGGGGACCTT CTAGAGGTTT CCTTGACTGA TCCCGACCTC AACTTGTATA	1597
cons ADN 41,42,43	-----	438
Consensus	AGGGGACCTT CTAGAGGTTT CCTTGACTGA TCCCGACCTC AACTTGTATA	1600

68/69

## FIG 52 B

Trans of MSRV pol cons prot 41,42,43	IMPESPTPLL GRDILAKAGA IHLNIGKGI PICCPLEEGB INPEVWAIIEG	50
Consensus	.....	50
Trans of MSRV pol cons prot 41,42,43	QYQOAKNARP VQVKLSDAS FPYQRYPLR PEALQGXQKI VKDLKAQGLV	100
Consensus	.....	100
Trans of MSRV pol cons prot 41,42,43	KPCSSPCNTP ILGVRKPNQO WRLVQDLRII NEAVFPLYPA VSSPYTLLSL	150
Consensus	.....	150
Trans of MSRV pol cons prot 41,42,43	IPEEAWEFTV LDLKDAFFCI PVRPDSQFLF AFEDPLNPTS QLTWIVLPQG	200
Consensus	.....	200
Trans of MSRV pol cons prot 41,42,43	FRDSPHLFGQ ALADLSQSS YLDTLVLYV DDLLIWARSE TLCHQATQST	250
Consensus	.....	36
Consensus	.....	250
Trans of MSRV pol cons prot 41,42,43	LIIFLTTCGYK VSKPKALCS QEIFYLGLKL SKGTRALSEE RIQPIIAYPH	300
Consensus	.....	86
Consensus	.....	300
Trans of MSRV pol cons prot 41,42,43	PKTLKQLRGE LGITGFQRKQ IPRYTHIARP LMTITRETQK ANIMLVWRTP	350
Consensus	.....	136
Consensus	.....	350
Trans of MSRV pol cons prot 41,42,43	TEVAFQALKK ALTQAPVFSL PTGQDFSLYA TEKIGIALGV LTQVSGMSLQ	400
Consensus	.....	136
Consensus	.....	400
Trans of MSRV pol cons prot 41,42,43	PVVYLSNED VVAKGWPHCL WVMAAVAVLV SEAVKIIQGR DLTWVTSNDV	450
Consensus	.....	140
Consensus	.....	450
Trans of MSRV pol cons prot 41,42,43	NGILTAKGDL WLSINHLINQ QALLLEEFVL RLRTCATLKP ATFLPDNEEK	500
Consensus	.....	146
Consensus	.....	500
Trans of MSRV pol cons prot 41,42,43	IEHNCQQVIA QTYAARGDLL EVPLTDPDLN LYTDGSSLAE KGLRKAGYAV	550
Consensus	.....	146
Consensus	.....	550
Trans of MSRV pol cons prot 41,42,43	ISDNGILESN RLTPGTSABL AELIALTWAL ELGEGKRVNI YSDSKYAYLV	600
Consensus	.....	146
Consensus	.....	600
Trans of MSRV pol cons prot 41,42,43	LHAHAIIWRE REFLTSEGTP INHQEAIRRL LLAVQKPKEV AVLHCQGHQE	650
Consensus	.....	146
Consensus	.....	650
Trans of MSRV pol cons prot 41,42,43	EEEREIEGMR QADIEAKKAA RQDSPLEMLI EGP	683
Consensus	.....	146
Consensus	.....	683

## FIG 53A

cons ADN 41,42,43	GACTTGAGCC AGTCATCATA CCTGGACAT CTGTGTCATC GGTACATGGA	50
cons ADN 1,5,8	GACTTGAGCC AGTCATCATA CCTGGACAT CTGTGTCATC AGTATRCGGA	50
Consensus	GACTTGAGCC AGTCATCATA CCTGGACAT CTGTGTCATC GGTATRCGGA	50
cons ADN 41,42,43	TGAMTTATTT TAGCCACCC ATTGAGAAAC CTGTGTCAT	100
cons ADN 1,5,8	TGAMTTATTT TAGCCACCC ATTGAGAAAC CTGTGTCAT	100
Consensus	TGAMTTATTT TAGCCACCC ATTGAGAAAC CTGTGTCAT	100
cons ADN 41,42,43	AAGCTCTCTT AAATTTCCTT GCTACCTGTG GGTACAAGGT TTCCAAACCA	150
cons ADN 1,5,8	AAGCTCTCTT AAATTTCCTT GCTACCTGTG GGTACAAGGT TTCCAAACCA	141
Consensus	AAGCTCTCTT AAATTTCCTT GCTACCTGTG GGTACAAGGT TTCCAAACCA	150
cons ADN 41,42,43	AGGGCTCASC TCTGCTCACA SCAGGTAA TACTTAGGGC TAAAATTATC	200
cons ADN 1,5,8	AGGGCTCASC TCTGCTCACA SCAGGTAA TACTTAGGGC TAAAATTATC	191
Consensus	AGGGCTCASC TCTGCTCACA SCAGGTAA TACTTAGGGC TAAAATTATC	200
cons ADN 41,42,43	CAAAGCTCC AGGCTCCTCA GAGAGGAACG TATCCAGCT ATACTGGMTT	250
cons ADN 1,5,8	CAAAGCTCC AGGCTCCTCA GAGAGGAACG TATCCAGCT ATACTGGMTT	241
Consensus	CAAAGCTCC AGGCTCCTCA GAGAGGAACG TATCCAGCT ATACTGGMTT	250
cons ADN 41,42,43	ATCCCATCC CAAACCTTA AAGCAACTAA GAGGTTCCT TGGCATAACA	300
cons ADN 1,5,8	ATCCCATCC CAAACCTTA AAGCAACTAA GAGGTTCCT TGGCATAACA	291
Consensus	ATCCCATCC CAAACCTTA AAGCAACTAA GAGGTTCCT TGGCATAACA	300
cons ADN 41,42,43	GGTTCCTGCC AATATGGAT TCCCGCTAC AGCAARFAG CCAGGCCATT	350
cons ADN 1,5,8	GGTTCCTGCC AATATGGAT TCCCGCTAC AGCAARFAG CCAGGCCATT	341
Consensus	GGTTCCTGCC AATATGGAT TCCCGCTAC AGCAARFAG CCAGGCCATT	350
cons ADN 41,42,43	AAATACACGA ATTAAGGAAA CTCAGAAAGC CATTACCCAT TAGTAAGAT	400
cons ADN 1,5,8	AAATACACGA ATTAAGGAAA CTCAGAAAGC CATTACCCAT TAGTAAGAT	391
Consensus	AAATACACGA ATTAAGGAAA CTCAGAAAGC CATTACCCAT TAGTAAGAT	400
cons ADN 41,42,43	GGACACTGA ACAGAAAGTG GCTTTCAGG CCCTAAAG	438
cons ADN 1,5,8	GGACACTGA ACAGAAAGTG GCTTTCAGG CCCTAAAG	429
Consensus	GGACACTGA ACAGAAAGTG GCTTTCAGG CCCTAAAG	438

## FIG 53B

cons prot 41,42,43	DLSQSSYLD LVLVYDDLL IATHSETLH QATQALLNFL ATCGKVSKE	50
cons prot 1,5,8	DLSQSSYLD LVLVYDDLL IATHSETLH QATQALLNFL ATCGKVSKE	47
Consensus	DLSQSSYLD LVLVYDDLL IATHSETLH QATQALLNFL ATCGKVSKE	50
cons prot 41,42,43	KAQLCSQVVK YLGLKLSKT RILHEERIQ ILVPHPKTL KQLTFLGIT	100
cons prot 1,5,8	KAQLCSQVVK YLGLKLSKT RILHEERIQ ILVPHPKTL KQLTFLGIT	97
Consensus	KAQLCSQVVK YLGLKLSKT RILHEERIQ ILVPHPKTL KQLTFLGIT	100
cons prot 41,42,43	AFCTIWIPRY SRIARPLNR IKETQKANTH IWRWTPPEEV AFQALK	146
cons prot 1,5,8	AFCTIWIPRY SRIARPLNR IKETQKANTH IWRWTPPEEV AFQALK	143
Consensus	AFCTIWIPRY SRIARPLNR IKETQKANTH IWRWTPPEEV AFQALK	146



# INTERNATIONAL SEARCH REPORT

International Application No

PCT/IB 97/01482

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/48 C12N5/08 C12N7/02 C07K14/15 C12N9/12  
C12N9/22 C12Q1/70 C07K16/10 G01N33/569 A61K39/21  
A61K39/42 A61K48/00

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12Q C07K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP 0 731 168 A (BIO MERIEUX) 11 September 1996 see the whole document ---	1-35
A	WO 95 21256 A (BIO MERIEUX ; PERRON HERVE (FR); MALLET FRANCOIS (FR); MANDRAND BER) 10 August 1995 see the whole document ---	1-35
A	WO 94 28138 A (UNIV LONDON ; GARSON JEREMY (GB); TUKE PHILIP (GB)) 8 December 1994 see the whole document ---	1-35
	--- -/--	



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

### \* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

22 April 1998

Date of mailing of the international search report

08/05/1998

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Hagenmaier, S

## INTERNATIONAL SEARCH REPORT

Inter. Patent Application No.

PCT/IB 97/01482

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	PERRON H ET AL: "IN VITRO TRANSMISSION AND ANTIGENICITY OF A RETROVIRUS ISOLATED FROM A MULTIPLE SCLEROSIS PATIENT" RESEARCH IN VIROLOGY, vol. 143, no. 5, 1 January 1992, pages 337-350, XP000569296 see the whole document ---	1-35
P, X	PERRON ET AL.: "MOLECULAR IDENTIFICATION OF A NOVEL RETROVIRUS REPEATEDLY ISOLATED FROM PATIENTS WITH MULTIPLE SCLEROSIS" PNAS, vol. 94, July 1997, pages 7583-7588, XP002062853 see the whole document ---	1-35
P, A	WO 97 06260 A (BIO MERIEUX ;PERRON HERVE (FR); BESEME FREDERIC (FR); BEDIN FREDER) 20 February 1997 see the whole document -----	1-35

# INTERNATIONAL SEARCH REPORT

information on patent family members

International Application No

PCT/IB 97/01482

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0731168 A	11-09-96	FR 2731356 A	13-09-96
		AU 5007396 A	02-10-96
		BR 9605926 A	02-09-97
		CA 2171242 A	10-09-96
		CZ 9603287 A	12-03-97
		WO 9628552 A	19-09-96
		JP 8322579 A	10-12-96
		NO 964760 A	08-11-96
		PL 317200 A	17-03-97
WO 9521256 A	10-08-95	FR 2715936 A	11-08-95
		FR 2715938 A	11-08-95
		FR 2715939 A	11-08-95
		FR 2715937 A	11-08-95
		FR 2727428 A	31-05-96
		FR 2728585 A	28-06-96
		AU 1711495 A	21-08-95
		CA 2141907 A	05-08-95
		DE 674004 T	19-09-96
		EP 0674004 A	27-09-95
		FI 954699 A	03-10-95
		JP 8511170 T	26-11-96
		NO 953925 A	04-12-95
WO 9428138 A	08-12-94	AU 676568 B	13-03-97
		AU 6760094 A	20-12-94
		CA 2163641 A	08-12-94
		EP 0700441 A	13-03-96
		JP 8511936 T	17-12-96
WO 9706260 A	20-02-97	FR 2737500 A	07-02-97
		AU 6823296 A	05-03-97
		EP 0789077 A	13-08-97
		NO 971493 A	03-06-97
		PL 319512 A	18-08-97

